

# White Paper Report

Report ID: 107667

Application Number: HT-50069-12

Project Director: Tanya Clement (TCLEMENT@ISCHOOL.UTEXAS.EDU)

Institution: University of Texas, Austin

Reporting Period: 9/1/2012-9/30/2015

Report Due: 12/31/2015

Date Submitted: 1/28/2016

**White Paper: High Performance Sound Technologies for Access and Scholarship**

HT-50069-12

High Performance Sound Technologies for Access and Scholarship (HiPSTAS)

PI: Tanya Clement

CoPI: Loretta Auvil

CoPI: David Tcheng

Office of Digital Humanities

January 25, 2016

## 1. Background

In August 2010, the Council on Library and Information Resources (CLIR) and the Library of Congress (LoC) issued a report titled *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age* (2010). This report argues that our sound heritage continues to deteriorate on legacy formats making digitization of the utmost importance, but that preservation and access cannot be solved through digitization alone. As Mark Greene and Dennis Meissner remark in their 2005 article “More Product, Less Process,” “processing is not keeping up with acquisitions, and has not been for decades, resulting in massive backlogs of inaccessible collections at repositories across the country” (p. 208). The same is true of unprocessed and therefore inaccessible archives of spoken word sound collections that hold important cultural artifacts.

The 2010 report suggests that if scholars and students do not use sound archives, our cultural heritage institutions will be less inclined to preserve them. Consequently, archives and libraries must collaborate with patrons and scholars to understand how recordings are and might be used. If librarians and archivists need to know what scholars and students want to do with sound artifacts in order to make these collections more accessible, then humanities scholars, arguably, also need to know what kinds of analysis are possible in an age of large, freely available collections and advanced computational analysis.

Software development for accessing and analyzing audio in general is underdeveloped, however. Observing the general dead air in this audio access and preservation soundscape, CLIR’s *Survey of the State of Audio Collections in Academic Libraries* (Smith, et. al, 2004) and CLIR’s report with the LoC, *National Recording Preservation Plan* (Nelson-Strauss, et. al, 2012), cite copyright legislation reform, organizational initiatives for shared preservation networks, and improvements in the processes of discovery and cataloging as the areas where research and development are most needed for increasing access. They call for “new technologies for audio capture and automatic metadata extraction” (Smith, et. al, 2004, 11) with a “focus on developing, testing, and enhancing science-based approaches to all areas that affect audio preservation” (Nelson-Strauss, et. al, 2012, 15) to help relieve these dark backlogs of undescribed, even though digitized, audio collections.

Currently, there are a few free open-source content management systems that enhance access to audio and video in well-designed environments. For example, both the Avalon Media System at Indiana and Northwestern and the Oral History Metadata Synchronization project (OHMS) out of the University of Kentucky are open source systems specifically designed for managing large collections of digital audio and video files. Avalon and OHMS enable users to curate, distribute and provide online access to their collections for purposes of teaching, learning and research.

Other systems allow users to segment or to create clips and playlists for audio organization such as the Stories Matter Project at Concordia. Finally, proprietary tools such as Kairos and Glifos are content management tools that help automate production and cataloguing on the backend while providing for rich media delivery over diverse data transport platforms and presentation devices. With these tools, varied materials that all relate to a single event can be linked and presented together, but these tools are designed to leverage human-generated texts and metadata. Systems that are dependent on such content do not scale well in a culture that is constantly producing large digital sound collections without transcripts or reliable metadata.

An opportunity for changing human-intensive practices of annotating and indexing sound seems present in the fact that computer performance—in terms of speed, storage capacity, and advancements in machine learning—has increased. The very popular Digging into Data Challenge, which is supported by funding agencies representing Canada, the Netherlands, the United Kingdom, and the United States is a testament to the wide array of perspectives and methodologies digital projects can encompass. The first (2009) and second (2011) rounds of awards include projects that are using machine learning and visualization to provide new methods of discovery. Some analyze image files (“Digging into Image Data to Answer Authorship Related Questions”) and the word in text files (“Mapping the Republic of Letters” and “Using Zotero and TAPoR on the Old Bailey Proceedings: Data Mining with Criminal Intent”). Other Digging into Data projects provide new methods for discovery with audio files by analyzing large amounts of music information such as the “Structural Analysis of Large Amounts of Music” (SALAMI) and “the Electronic Locator of Vertical Interval Successions (ELVIS)” projects or—large scale data analysis of natural language usage (the “Mining a Year of Speech” and the “Harvesting Speech Datasets for Linguistic Research on the Web” projects). None of these projects, however, take on the perspective of scholars and cultural heritage institutions focused on folk, literary, or tribal spoken word collections. The third round of this funding, announced in January 2014, does not include any sound projects at all.

At the same time, the work required for such inquiry is related to the work that scholars have been doing for decades on features of music. The Automated Learning Group (ALG) of the Illinois Informatics Institute (I3) (where ARLO has been developed), for example has years of experience in high-performance machine learning and audio analysis. J. Stephen Downie (PI of SALAMI) and Michael Welge developed a system for comparing different music information retrieval (MIR) systems. Downie and collaborators also developed NEMA (Networked Environment for Music Analysis), a multinational, multidisciplinary cyber-infrastructure project for music information processing that builds upon music information retrieval research. NEMA can be used for genre and mood classification as well as composer identification; for similarity retrieval where similarity is measured on prosodic features of pitch, tempo, and accent or the key or tone of music; and

structural segmentation evaluation that identifies the key structural sections in music such as a change in verse, movement, or the addition of a chorus. Scholarship produced by Downie and collaborators as a result of their interventions in music information retrieval (Downie, 2003, 2008; Downie, et al., 2010) was particularly helpful in guiding our infrastructure development in the HiPSTAS project.

## **2. Project Activities**

The HiPSTAS institute had two primary learning outcomes: (A) participants would produce new scholarship using audio collections with advanced technologies such as classification, clustering, and visualizations; and (B) participants would engage in the scholarly work of digital infrastructure development by contributing to recommendations for the implementation of a suite of tools for collecting institutions interested in supporting advanced digital scholarship in sound.

NEH support for HiPSTAS was used primarily to set up ARLO for testing with spoken word audio collections, to conduct two HiPSTAS meetings (including travel for participants and speakers), to support consulting and software maintenance over the course of the year for the participants, and to support project management and administrative activities.

### **Pre-Institute Activities**

In the first nine months of funding for HiPSTAS, we created an online presence for reporting and general information:

- about the project at <http://blogs.ischool.utexas.edu/hipstas/about/>
- Call for Participants at <http://blogs.ischool.utexas.edu/hipstas/institute/cfp/> [also attached in Appendix F.]
- participant project pages at <https://sites.google.com/site/nehhipstas/documentation>
- Notes from both virtual and in-person meetings:  
<https://sites.google.com/site/nehhipstas/archive>
- a discussion forum at <https://groups.google.com/forum/?fromgroups=#!forum/hipstas> .

Clement and the UT GRA developed, advertised, and implemented an online application process for participants (shown in the link above). Our advisory board subsequently vetted and chose 20 participants, including humanities junior and senior faculty and advanced graduate students as well as librarians and archivists from across the U.S. who are interested in developing and using new technologies to access and analyze spoken word recordings within audio collections.

Our twenty participants came from a wide range of types of institutions and professional backgrounds: • Jeffrey Boruszak: PhD Candidate, Department of English, University of Texas-Austin • Kim Christen: Associate Professor and Director of Digital Projects, Washington State University • Hartwell Francis: Assistant Professor, Anthropology and Sociology, Western Carolina University • Michael Hennessey: Editor, PennSound and Jacket2 • Jennifer Himmelreich: MLIS Student, San Jose State University • Kira B. Homo: Electronic Records Archivist, University of Oregon • Justin Kovar: Digitization Project Archivist, The Dolph Briscoe Center For American History • Michael J. Kramer: Lecturer, Northwestern University • Bert Lyons: Digital Archivist, Library of Congress • Stephen McLaughlin: Director, PennSound Radio; Podcasts Editor, Jacket2.org • Ben Miller: Assistant Professor, Department of English, Georgia State University • Virginia Millington: Recording and Archive Manager, StoryCorps • Michael Nardone: PhD Candidate, Concordia University • Linda Newman: Digital Projects Coordinator, Langsam Library, University of Cincinnati • Juliana Nykolaiszyn: Assistant Professor; Oral History Librarian, Oklahoma State University • Amber Paraniek: Librarian, Library of Congress • Gena Peone: Assistant Cultural Collections Manager, Cultural Preservation Department, Spokane Tribe of Indians • Eric Rettberg: Postdoctoral Preceptor, University of Virginia • Elizabeth Russey Roke: Digital Archivist, Emory University • Kristen Suagee-Beauduy: Graduate Assistant, Department of English, Western Carolina University • Kenneth Sherwood: Associate Professor of English; Co-Director of Center for Digital Humanities and Culture, Indiana University of PA • Dustin Tahmahkera: Assistant Professor, Department of Communication Studies, Humanities Division, Southwestern University.

The participants' previous experience with audio analysis had been limited to the textual content of the audio files or, in some cases, waveform representations. The applications they were using (such as Audacity and iTunes) provided them with very limited access to the collections' sound features, features that Charles Bernstein and others have identified as significant for analysis. Bernstein (2011) claims that waveforms can only identify part of what makes poetry audio files interesting. He explains that "[t]here are four features or vocal gestures, that are available on tape but not page that are of special significance for poetry;" these include "the cluster of rhythm and tempo (including word duration), the cluster of pitch and intonation (including amplitude), timbre, and accent" (p. 144). The features that signify meaning in sound are diverse and a waveform can only visualize part of the first cluster (tempo) and part of the second cluster (amplitude). As such, though they were accustomed to working daily with large spoken text audio collections, prior to the HiPSTAS Institute, participants had never had access to the types of sound features machine-learning tools rely on for analyzing audio.

During the Institute, participants were introduced to ARLO (Adaptive Recognition with Layered Optimization), a machine learning application for analyzing large sound collections that was

originally developed to classify and analyze bird calls by extracting audio features and displaying the audio data as a spectral graph (Tcheng, et al., 2009 and Enstrom, et al., 2008). ARLO was then extended with National Science Foundation (NSF) funding to help another set of scholars classify pollen grains using image features instead of audio features (Tcheng, et al., 2008). ARLO extracts basic prosodic features such as pitch, rhythm and timbre for visualizations and spectral matching. These features are then used for classification and clustering. HiPSTAS shows that ARLO, which has been proven effective in the sciences, can also be used as a supercomputing resource for analyzing patterns across spoken word collections that are of interest to humanists.

Clement was approved through IRB at UT Austin to conduct usability studies with the HiPSTAS participants, with whom she conducted and recorded hour-long interviews before the Institute asking questions about their practices with access and scholarship with spoken word collections (see Appendix C for IRB application and interview script). Through subsequent interviews, observations and surveys, Clement gathered feedback throughout the HiPSTAS Institute in order to help inform development and evaluation.

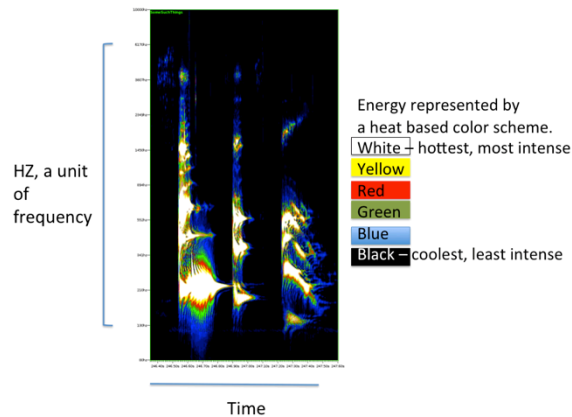
#### *ARLO Software Development*

Participants primarily used the ARLO (Adaptive Recognition with Layered Optimization) software to analyze sound files since the I3 team saw its development as more advantageous to the project than developing NEMA, which was no longer under development at the start of this project.

ARLO, which was originally called NESTER, was developed with UIUC seed funding for avian ecologist David Enstrom (2008) to begin exploring the use of machine learning for data analysis in the fields of animal behavior and ecology. Specifically, by extracting basic prosodic features such as pitch, rhythm and timbre for classification, clustering, and visualizations, Enstrom used ARLO to identify and catalog all syllables (phonemes) produced by a species of songbird, the Northern Cardinal. With over 2400 hours of recordings, Enstrom used the pattern recognition and machine learning functions of ARLO to automatically produce bird vocalization transcripts and analyze vocal patterns in the audio data streams. As such, ARLO allowed Enstrom to process these streams and test hypotheses regarding song production and culture in birds that have heretofore been intractable.

ARLO analyzes event wave forms (raw audio samples) by extracting time and frequency information in the form of a spectrogram (see Figure 1). The spectrogram is computed using band pass filters linked with energy detectors. For example, in the spectrogram shown in Figure 1, the color of each pixel represents the numerical value of energy (which represents the sum of potential and kinetic energy) of a particular frequency at a point in time. Using a heat based color scheme,

the lowest values are black (cool), blue, green, red, yellow, and the points with the highest or most intense energy values are white. ARLO spectrograms contain similar information as FFT based spectrograms and are flexible in application because the frequencies and damping factors can be optimized for a given problem.



**Figure 1: A spectrogram produced by ARLO of Gertrude Stein saying “some such thing” from a reading of her novel *The Making of Americans*.**

The machine-learning algorithm ARLO uses for classification to find events in audio is called “instance based learning” (IBL). In IBL, the machine memorizes a number of classified training examples and matches them against new unseen examples to predict events. In ARLO, examples are audio events defined by a start and end time such as a two-second clip. ARLO finds matches by taking each known classified example and “sliding” it across new audio files looking for good matches. The number of match positions considered per second is adjustable and is set to the spectra sample rate. The degree of match or “match strength” is measured by the correlation between spectrograms, which can range between -1.0 and +1.0. In addition to simple spectra matching, a user can isolate pitch and volume traces, compute correlations on them, and weight the different feature types when computing the overall match strength. This allows the user to preferentially weight spectral information that might correspond to such aspects as pitch or rhythm.

For clustering, ARLO uses a single-threaded algorithm, ARLO selects a randomly chosen subset of examples then divides these examples into piles or clusters of equal or unequal size. ARLO evaluates the quality of every cluster by first computing the cluster centroid as an average of all the spectra and then averaging the distance of the examples in each cluster to the computed centroid. Finally, ARLO iterates over the following steps a given number of times:

1. randomly select two clusters and an example from each cluster to evaluate;



2. compute the cohesiveness of these two clusters based on exchanging the selected examples;
3. if the cohesiveness of the clustering improves, then exchange these two examples

Developing ARLO as a web-based application that can be leveraged by a wide community of users has been an essential goal of the HiPSTAS project. Before the HiPSTAS project, the ARLO front-end (Python/Django web interface) and back-end (Java processes and MySQL database) ran on a single dedicated server hosted by the National Center for Supercomputing Applications (NCSA) at UIUC. As a community tool, we have implemented ARLO on Stampede, a National Science Foundation funded petascale High Performance Computing (HPC) system at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. All users and projects are on the same machine and each user can create projects to keep their data private or share projects with each other. Having a single installation minimized maintenance and deployment times during development, but the Stampede server can handle the workload from user interactions on the front-end for numerous users at once.

Because HiPSTAS participants were able to use the ARLO implementation on Stampede, we were able to gather user and technical requirements for further developing ARLO for super computing tasks of interest to humanists working with spoken word sound collections. When describing the inherent difficulties in developing music information retrieval systems, Downie (2008) and Downie, et al. (2010) identify similar issues we faced when developing ARLO as a community tool. Downie identifies ten major research issues that must be addressed when developing MIRs including determining effective procedures and evaluation techniques for (1) indexing; (2) retrieval queries; (3) user interface design for access and analysis; (4) audio compression for efficient processing; (5) audio feature detection that yields productive analyses; (6) machine learning algorithms; (7) classification techniques; (8) security measures for sensitive materials; (9) accessibility procedures for a range of user communities; and (10) sufficient computing and storage infrastructure development for data-intensive techniques (Downie, 2008; Downie, et al., 2010). As described in the next section, our concern thus far in HiPSTAS has been focused on gathering user requirements for determining audio features that yield productive queries and infrastructure requirements that support appropriate computing and storage needs.

This development work for HiPSTAS included limited interface development with ARLO for non-birding humanities users, such as the ability to analyze longer files, adding short keys for play, stop, fast-forward, etc. This minor interface development allowed humanities users to test the machine learning system and perform exploratory discovery (clustering) and automated classification (prediction or supervised learning) processes as well as visualizations. Further, infrastructure development that allowed multiple users to use multiple collections and create

separate tags or annotation sets and share them. Finally, the project consultant, Tony Borries, Clement, Tcheng, and the UT GRA spent significant time on creating documentation for ARLO including:

- Documentation specifically designed for humanists interested in analyzing spoken text audio collections including step-by-step instructions as well as instructional videos: <https://sites.google.com/site/nehhipstas/documentation>.
- Documentation for developers: [http://wiki.arloproject.com/Main\\_Page](http://wiki.arloproject.com/Main_Page)

The original implementation of ARLO for modeling ran in parallel on systems at the National Center for Supercomputing Applications (NCSA). As part of HiPSTAS, the ARLO backend (written in Java) was developed to make calls to the Texas Advanced Computing Center's Stampede system.

For the HiPSTAS Institute, infrastructure development work was necessary in order to implement the ARLO backend on the Stampede system. This development included developing task scheduling technology, separating out and distributing Java processes, and setting up community user accounts for job batching and task management. Transitioning ARLO, which was implemented on a single processor, to achieve parallelism on Stampede necessitated developing a task manager to create sub problems and a task handler to do the work and return the results.

For the Institutes, we ingested 27,000 files from PennSound and 150 hours of folklore from the Dolph Briscoe Center for American History and set up user accounts for the HiPSTAS participants. Hooked up to Stampede, ARLO could accomplish computational tasks that required more processing power such as finding patterns across 27,000 PennSound audio files. Reading this amount of data, without precomputing or indexing would have taken days on a regular system, which would have precluded our ability to implement a sandbox for humanities scholars at the Austin-based Institutes.

### **First HiPSTAS Institute, May 29 – June 1, 2013**

The first four-day meeting of HiPSTAS held met at the University of Texas, May 29 – June 1, 2013, participants were introduced to essential issues that archivists, librarians, humanities scholars, and computer scientists and technologists face in understanding the nature of digital sound scholarship and the possibilities of building an infrastructure for enabling such scholarship. The Co-PI's developed a workshop to introduce participants to advanced computational analytics such as clustering, classification, and visualizations with ARLO (<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/first-meeting-may-29-june-1-2013/>). See **Appendix A** for first meeting agenda.

In the first year, Clement was also a significant spokesperson for the project, giving multiple talks about the project at various venues including scholarly conferences, universities and more popular venues:

*Invited Talks and Panels (first year, by Tanya Clement)*

- Plenary, “‘This is Just to Say’: Changing the Nature of Poetry Performance Studies and Learning to Listen with Machines,” University of Cincinnati (November 2013).
- Plenary, “Project-based DH in the Undergraduate Curriculum: a History, a Few Principles, and Some Suggestions” Arthur Vining Davis Digital Humanities Summer Faculty Workshop, Northwestern University (August 2013).
- Plenary, “Poetry in Motion: Using High Performance Sound Technologies for Access and Scholarship with Spoken-Word Collections” Poetry vs. Philosophy: Life, Artifact & Theory Symposium, Texas A&M University (April 2013).
- Invited talk, “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” University of Victoria, British Columbia, Canada (March 2013).
- Invited talk, “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” Approaching The Poetry Series: Using Literary Recordings as Scholars and Digital Designers, Concordia University, Montreal, Quebec, Canada (March 2013).
- Keynote: “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” Nebraska Forum on Digital Humanities “Hacking at Books”, Center for Digital Research in the Humanities. University of Nebraska – Lincoln (February 2013).

*Refereed Conference Panels and Presentations (first year, by Tanya Clement)*

- “What We Talk About When We Talk About Sound.” Digital Frontiers Conference, University of North Texas (September 2013).
- “Sound Seeings or High Performance Sound Technologies for Access and Scholarship” SXSWInteractive, Austin, TX (March 2013).

**Virtual Meetings, Interim year, June 2013 to May 2014**

Public virtual meetings (notes: <https://sites.google.com/site/nehhipstas/archive/virtual-meetings>) continued monthly over the course of the year with project team members and participants calling in.

Over the course of the year, the HiPSTAS teamwork yielded three significant results for the computational analysis of spoken word collections of keen interest to the humanities: (1) an assessment of user requirements; (2) an assessment of technological infrastructure needed to support a community tool; (3) preliminary experiments using these advanced resources that show the efficacy, both in terms of user needs and computational resources required, of using machine learning tools to improve discovery with unprocessed audio collections [these experiments and results appear in Clement et. al, 2014.]

## **Second HiPSTAS Institute, May 29 – June 1, 2014.**

In the second Institute, project team members discussed the evolution of the project and participants presented their work in a public event on the first day. On the second day, participants met to discuss recommendations for the continued development of High Performance Sound Technologies for Access and Scholarship. See **Appendix B** for second meeting agenda.

In the second year, Clement continued to speak for the project, giving multiple talks about the project at various venues including scholarly conferences, universities and more popular venues:

### *Invited Talks and Panels (second year, by Tanya Clement)*

- Invited talk, “The Aporia of Sound: Digital Humanities Futures in Poetry Performance Studies” Cornell University. The Conversations in Digital Humanities speaker series. Cornell University, Ithaca, NY (November 2015).
- Invited talk, “Part of Algorithmic Audition: A Roundtable on DH and Sound. The Center for Digital Humanities @ Princeton, Princeton University, Princeton, New Jersey (November 2015).
- Invited talk, “High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Text Mining In The Humanities: Collaboration across Sectors, University of Ottawa, May 31, 2015.
- Invited talk, “High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Digital Asset Symposium (DAS) (May 2015).
- Invited talk, “HiPSTAS: High Performance Sound Technologies for Access and Scholarship,” Cultural Analytics Conference, University of Chicago (May 2015).
- Invited talk, “Towards a Rationale of Audio Text: Audio Exploration and High Performance Technologies,” Penn Digital Humanities Forum, University of Pennsylvania (April 2015).
- Keynote, “The Shunting Yard: Developing Infrastructures for Meaning Making in Information and Sound Studies.” Texas Digital Humanities Conference, University of Houston (April 2014).
- Invited talk, “What We Talk about When We Talk about Sound: Introducing High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Sound + Conference, Center for Literary and Comparative and Studies of the English Department, University of Maryland, College Park (March 2014).
- Invited talk, “HiPSTAS, What?: Information Retrieval, Machine Learning, and Visualizations with Sound,” CUNY Digital Humanities Initiative, New York (March 2014).
- Invited talk, “Introducing High Performance Sound Technologies for Access and Scholarship,” Columbia University, New York (March 2014).

### *Refereed Conference Panels and Presentations (second year, by Tanya Clement)*

- “Machinic Ballads: Alan Lomax’s Global Jukebox and the Categorization of Sound Culture” Society for Ethnomusicology Annual Conference, Austin, TX, (December 2015).
- “Reconsidering the Hermeneutics of Listening with High Performance Sound Technologies for Access and Scholarship (HiPSTAS). Inertia: A Conference on Sound, Media, and the Digital Humanities, University of California, Los Angeles (May 2015).
- “Do you Hear What I Hear? Introducing High Performance Sound Technologies for Access and Scholarship.” Society of American Archivists Annual Meeting, Washington DC (August

- 2014).
- “Developing for Distant Listening: Developing Computational Tools for Sound Analysis By Framing User Requirements within Critical Theories for Sound Studies” Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “<audio>Digital Humanities</audio>: The Intersections of Sound and Method.” Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “The Hermeneutics of Distant Listening to Spoken Word Texts with High Performance Sound Technologies for Access and Scholarship” AV in DH, 2014 Workshop, Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “Using Sound Technologies in the Study and Documentation of Spoken Word Recordings.” Native American and Indigenous Studies (NAISA) Annual Meeting, Austin, TX (May 2014).

*Publications by and about HiPSTAS:*

- Clement, T., McLaughlin, S., Tchong, D., Auvil, L., Borries, T. “Measured Applause: Toward a Cultural Analysis of Audio Collections” *Cultural Analytics* (under review).
- Clement, T. “A Rationale of Audio Text.” *digital humanities quarterly* (under review).
- Clement, T. “The Ear and the Shunting Yard: Meaning Making as Resonance in Early Information Theory.” *Information & Culture* 49.4 (2014): 401-426.
- Clement, T. “Word. Spoken. Articulating the Voice as Descriptive Metadata for High Performance Sound Technologies for Access and Scholarship (HiPSTAS).” In *Provoke*. Darren Mueller, Mary Caton Lingold and Whitney Anne Trettien (eds.) Durham, NC: Duke University Press (Forthcoming).
- Clement, T. “Introducing High Performance Sound Technologies for Access and Scholarship.” *The International Association of Sound and Audiovisual Archives Journal* (September 2013) 41: 21-28.
- Clement, T. “When Texts of Study are Audio Files: Digital Tools for Sound Studies in DH” In Susan Schreibman, Ray Siemens and John Unsworth (eds.), *A New Companion to Digital Humanities* (Blackwell Companions to Literature and Culture), 2016: 348-357.
- Clement, T. “Machinic Ballads: Alan Lomax’s Global Jukebox and the Categorization of Sound Culture.” *Sounding Out! Blog*. 14 May 2015.
- Clement, T., Tchong, D., Auvil, L., and Borries, T. “High Performance Sound Technologies for Access and Scholarship (HiPSTAS) in the Digital Humanities” Proceedings of the 77th Annual ASIST Conference, Seattle, WA, 31 October – 5 November.
- Clement, T. and Roy, L., “HiPSTAS: An Institute Advancing Tools for Analyzing Digital Audio Collections,” *American Indian Library Association Newsletter* 36 (2013): 8-15.
- Filreis, A. “Anti-ordination in the visualization of the poem's sound” *Jacket2*. 27 February 2014. <http://jacket2.org/commentary/anti-finality-visualization-poems-sound>
- Francis\*, H., Clement, T., Peone, G., Carpenter, B., Suagee-Beauduy, K. “Accessing Sound at Libraries, Archives, and Museums” *Indigenous Ownership & Libraries, Archives, and Museums* (forthcoming).
- MacArthur, M. “Monotony, the Churches of Poetry Reading, and Sound Studies.” *PMLA*. Forthcoming 2016.
- Mustazza, C. ” “The noise is the content: Toward computationally determining the provenance of poetry recordings" using ARLO. *Jacket2*. 10 Jan. 2015. <https://jacket2.org/commentary/noise-content-toward-computationally-determining-provenance-poetry-recordings>
- Perez-Hernandez, D. “Scholars Collaborate to Make Sound Recordings More Accessible” *The Chronicle of Higher Education*. 26 March 2014. <http://chronicle.com/blogs/wiredcampus/scholars-collaborate-to-make-sound-recordings-more-accessible/51215>
- Rettberg, E. “Hearing the Audience”. *Jacket2*. 26 March 2015.

- <http://jacket2.org/commentary/hearing-audience>
- Sherwood, K. “Distanced sounding: ARLO as a tool for the analysis and visualization of versioning phenomena within poetry audio” *Jacket2*. 2 March 2015.  
<https://jacket2.org/commentary/distanced-sounding-arlo-tool-analysis-and-visualization-versioning-phenomena-within-poetr>

#### 4. Accomplishments

We planned several important results from the HiPSTAS Institute. The curriculum of this Institute is premised on the idea that building scholarly infrastructure in the digital humanities is the work of scholars, librarians and archivists, and computer scientists together.

Our resulting scholarship (listed above) reflects deep collaborations across scholars, computer scientists and cultural heritage professionals and a knowledge of digital sound preservation and computational analysis of sound. Further, the participating collections have increased their ability to allow users to perform new kinds of scholarship with the data sets we created during the grant period. Finally, the recommendations we have created for the development of advanced computational tools for digital scholarly inquiry in sound reflect the needs and concerns of both the stewards of sound collections and the scholars who use them. The publications listed above describe three use cases: (1) Poetry (Clement, “Word. Spoken.”; Clement, “When Texts of Study are Audio Files” and Clement, “Towards a Rationale of Audiotext”); (2) Folklore (Clement, “Machinic Ballads”); and (3) Recordings in Indigenous Communities (Francis\*, H., Clement, T., Peone, G., Carpenter, B., Suagee-Beauduy, K. “Accessing Sound at Libraries, Archives, and Museums”). The next NEH grant, described below, funds use case work we are currently doing with Archivists.

#### 5. Audiences

Specifically, the HiPSTAS Institute represented a wide range of professional communities that are necessarily impacted by increased access and scholarship with sound. Our 20 participants included:

- 9 librarians and archivists
- 8 humanities scholars
- 3 advanced graduate students in humanities and information science

These students, scholars, and practitioners represented interests in audio collections from diverse communities across the United States including Native American tribal communities and Civil Rights collections from the American South. The projects to which we had access during the Institute (namely, 30,000 audio files from PennSound and 57 feet of tapes (reels and audiocassettes) from the UT folklore collection at the Dolph the Briscoe Center for American History at UT Austin each represented a wide range of voices from across the United States. Further, participants came from communities that represent an even wider range of voices and communities. Collections associated with our participants include but are not limited to:

- Field recordings (200,000 recordings) American Folklife Center, Library of Congress
- 30,000 hours, Oral histories, Storycorps
- Speeches in the Southern Christian Leadership Conference recordings, Emory University
- 700 recordings in the Elliston Poetry Collection at the University of Cincinnati

- 36 interviews in the Dust, Drought and Dreams Gone Dry: Oklahoma Women and the Dust Bowl (WDB) oral history project out of the Oklahoma State Libraries

New audiences include many partners from a diverse range of disciplines (biology, humanities, linguistics, and libraries and archives); our Advisory Board which includes representatives who use and represent important AV archives and libraries such as the Archive of Indigenous Languages of Latin America (AILLA), the Macaulay Library of Natural Sounds, and the WGBH Media Library and Archives; and representatives from national aggregators such as DPLA, HTDL and the HTRC; as well as metadata and AV experts involved in the Europeana Sounds project, AVPreserve, CLIR, Pop Up Archive, and the UT iSchool.

## 6. Evaluation

The project was continually evaluated by project team members, project participants, and third party reviewers of presentations at conferences and scholarly publication venues as evidenced by the list of publications and presentations in Section 1: Project Activities. Evaluations were given to all the project participants at the second meeting Institute meeting in Austin, Texas in May, 2014. These evaluations were done in groups. The questions we asked included the following:

### *Group Evaluations*

1. In general, what are the ideal tools that would be most valuable to your community?
2. In particular, how can a tool like ARLO accommodate current workflow practices?
3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?
4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?
5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?
6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?
7. What are future grant or collaboration opportunities the group might engage to support this work?
8. What are future pedagogical or publishing opportunities the group might engage to support this work?
9. What aspect of the process of developing a tool like ARLO is/was most frustrating? How do we ameliorate
10. What are more questions we should be asking? (answer them, please)

### *Evaluation Results*

The results of the evaluation were that participants were enthusiastic about the continued development of ARLO and the use of such tools in their own research, in classrooms, with community members, and with collections management. The participants gave us copious response on what features were needed for further developing ARLO. These are included in Appendix D. More qualitative results for each of the questions above appear in Appendix E with four groups represented in these accounts: Archivists; Poetry Librarians and Scholars; Native American Librarians and Scholars; and Sound Researchers.

The project team's assessment of the program was very positive. It would have ultimately been useful to have more development money for ARLO, since it is not stable enough to share with scholars beyond the institute, but this has been the focus of subsequent grants that have taken advantage of both the content of these initial conversations as well as the relationships that were created.

#### *Public Response*

The public response was also very positive as evidenced by the Chronicle of Higher Education publication (cited above), as well as well-attended presentations at more popular venues such as SXSW and DAS. Further, because of her project leadership, Clement was asked to become a Research Associate for the Radio Preservation Task Force of the Library of Congress.

### **7. Continuation of the Project**

The success of the project is evidenced by the second grant we were awarded by the National Endowment for the Humanities Preservation and Access grant for "HiPSTAS Research and Development with Repositories" (HRDR). In this HRDR phase, we plan to leverage the conversations and collaborations we established as part of HiPSTAS to develop the ARLO software as a more generally accessible and usable tool for the wider humanities community at both small and large institutions and to teach these communities how to use it. To this end, the HRDR phase will include three primary products: (1) a release of ARLO (Automated Recognition with Layered Optimization) that leverages machine learning and visualizations to augment the creation of descriptive metadata for use with a variety of repositories (such as a MySQL database, Fedora, or CONTENTdm); (2) a workshop curriculum and documentation for wider dissemination and training with the software; and (4) a white paper that details best practices for automatically generating descriptive metadata for spoken word digital audio collections in the humanities.

HiPSTAS participant Marit MacArthur has received an ACLS digital innovation fellowship to develop the ARLO interface for humanists interested in pitch tracking. She is actively pursuing other granting possibilities.

The use of ARLO and development under Clement's direction plays a small part in the tools being developed in the recently IMLS-funded project involving WGBH and the Pop Up Archive for "Improving Access to Time-Based Media through Crowdsourcing and Machine Learning" (on which Clement is an Advisory Board member).

Finally, we have just applied for a third major grant through the Institute of Museum and Library Services (IMLS) for funds to implement and test implementing ARLO at the Texas Advanced Computing Center at the University of Texas. Like many tools developed for research, however, ARLO 1.0 lacks essential aspects for broader implementation such as user-tested interfaces and workflows that reflect the storage capacity and processing power needed to efficiently meet long-term demands for real users in a wide range of settings who want to access AV materials. This partnership includes many partners from a diverse range of disciplines (biology, humanities, linguistics, and libraries and archives); our Advisory Board which includes representatives who use and represent important AV archives and libraries such as the Archive of Indigenous Languages of Latin America (AILLA), the Macaulay Library of Natural Sounds, and the WGBH Media Library and Archives; and representatives from national aggregators such as DPLA, HTDL and the HTRC; as well as metadata and AV experts involved in the Europeana Sounds project, AVPreserve, CLIR, Pop Up Archive, and the UT iSchool.



To better understand and document the socio-technical needs associated with trying new, largescale, and machine-automated processes, our proposal seeks funding for completing the following deliverables:

- 1) A suite of tested interfaces for ARLO 2.0, including documentation, tutorials, and sample AV files.
- 2) A suite of tested machine learning algorithms for searching and identifying significant patterns in AV collections, including documentation, tutorials, and sample data sets.
- 3) A suite of API-driven executable code bases for implementing ARLO 2.0 on personal computers, local or cloud servers, and on supercomputer clusters, including documentation.
- 4) Reports and recommendations for implementing ARLO 2.0 at different scales for researchers from the varied fields represented by our use cases and for archives and libraries who seek to make their AV collections discoverable as part of the National Digital Platform. The reports will include sample workflows for mapping data generated by ARLO into DPLA Metadata Application Profile (Map) 4.0, the MARC21 format (which is used by HTDL), and PBCore, which is used by WGBH and the Pop Up Archive as well as recommendations for the creation of feature sets (as a means for offering information about collections while also preserving copyright and privacy restrictions)

## 8. Long Term Impact

Evidence of projects that have been inspired by the HiPSTAS Institute appear in the Native American, Archives, Poetry, and Sound Studies communities appear above in the list of publications. Currently, ARLO is not usable beyond a small set of people, which negatively impacts our ability to create a wider community of users and the impact this work might have in the classroom and in scholarship. At the same time, as the publications show, there is great interest. Anecdotal evidence also suggests that HiPSTAS participants have been applying for grant funding for their own sound projects.

## 9. Grant Products

We anticipated three primary outcomes of the HiPSTAS Institute, closely tied to evaluation, as having the most impact on advancing computational analysis in sound scholarship. These results have been achieved:

- a) **Publicly available ongoing evaluation on the process of scholarship and technical developments** including the website at the UT iSchool (<http://blogs.ischool.utexas.edu/hipstas/> and related sites linked there) that we have maintained as a public source for information about the project and a venue for disseminating final reports. Much of the dialog before the first Institute, in the interim year, and after the close of the Institute took place in project work spaces we created in free, open-source platforms on Google (<https://sites.google.com/site/nehhipstas/project-pages>). Beyond the monthly virtual meetings and the big mid-January 2014 meeting in Google Hangout, we have used Google Sites for an ongoing master document in which the project team has included documentation for Humanists seeking to use the ARLO software (<https://sites.google.com/site/nehhipstas/documentation>) as well as for developers ([http://wiki.arloproject.com/Main\\_Page](http://wiki.arloproject.com/Main_Page) ).
- b) **Curriculum and Scholarship:** The curriculum, including the ARLO labs, and the outcome of both meetings of the HiPSTAS Institute have been made openly available as part of the iSchool web site

(<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/first-meeting-may-29-june-1-2013/>) and (<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/second-meeting-may-27-28-2014/>). Subsequent to the final meeting, participants were invited to contribute scholarship to a special series of *Jacket2* magazine (a preeminent and open source venue for creative and scholarly, digital work) on experimental digital analyses of poetry audio, titled “Clippings” (these are listed in “Publications” above). The visibility of the curricular materials, and of participants’ samples and documentation via the web site and open source publication venues have made the results of this Institute accessible to a wider audience beyond those able to participate directly.

- c) **Final white paper and recommendations:** The final white paper, written by Clement, Tchong and Auvil, reflects the monthly status reports and the developing infrastructure in ARLO. This report includes recommendations for implementing an open-source, freely available suite of tools for supporting scholarship on audio files. The purpose of disseminating these recommendations through the UT iSchool is to offer best practices for cultural heritage institutes that are new to making their sound files available via Web services frameworks and to provide the final recommendations for developing and implementing a more robust technical infrastructure based on feedback about ARLO collected during the HipSTAS Institute. The report is being disseminated at <http://blogs.ischool.utexas.edu/hipstas/2016/01/25/hipstas-neh-institute-final-white-paper/>
- d) **ARLO code** (<https://bitbucket.org/arloproject/> )

## 10. Appendices

Appendix A: First Meeting Agenda, May 29 – June 1, 2013; University of Texas at Austin

Appendix B: Second Meeting Agenda, May 27 – 28, 2014; UT Austin iSchool,

Appendix C: IRB Application with Interview Protocol

Appendix D: ARLO Requirements for Humanists

Appendix E: Group Evaluations of ARLO

Appendix F: Call for Participants

## 1. Background

In August 2010, the Council on Library and Information Resources (CLIR) and the Library of Congress (LoC) issued a report titled *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age* (2010). This report argues that our sound heritage continues to deteriorate on legacy formats making digitization of the utmost importance, but that preservation and access cannot be solved through digitization alone. As Mark Greene and Dennis Meissner remark in their 2005 article “More Product, Less Process,” “processing is not keeping up with acquisitions, and has not been for decades, resulting in massive backlogs of inaccessible collections at repositories across the country” (p. 208). The same is true of unprocessed and therefore inaccessible archives of spoken word sound collections that hold important cultural artifacts.

The 2010 report suggests that if scholars and students do not use sound archives, our cultural heritage institutions will be less inclined to preserve them. Consequently, archives and libraries must collaborate with patrons and scholars to understand how recordings are and might be used. If librarians and archivists need to know what scholars and students want to do with sound artifacts in order to make these collections more accessible, then humanities scholars, arguably, also need to know what kinds of analysis are possible in an age of large, freely available collections and advanced computational analysis.

Software development for accessing and analyzing audio in general is underdeveloped, however. Observing the general dead air in this audio access and preservation soundscape, CLIR’s *Survey of the State of Audio Collections in Academic Libraries* (Smith, et. al, 2004) and CLIR’s report with the LoC, *National Recording Preservation Plan* (Nelson-Strauss, et. al, 2012), cite copyright legislation reform, organizational initiatives for shared preservation networks, and improvements in the processes of discovery and cataloging as the areas where research and development are most needed for increasing access. They call for “new technologies for audio capture and automatic metadata extraction” (Smith, et. al, 2004, 11) with a “focus on developing, testing, and enhancing science-based approaches to all areas that affect audio preservation” (Nelson-Strauss, et. al, 2012, 15) to help relieve these dark backlogs of undescribed, even though digitized, audio collections.

Currently, there are a few free open-source content management systems that enhance access to audio and video in well-designed environments. For example, both the Avalon Media System at Indiana and Northwestern and the Oral History Metadata Synchronization project (OHMS) out of the University of Kentucky are open source systems specifically designed for managing large collections of digital audio and video files. Avalon and OHMS enable users to curate, distribute and provide online access to their collections for purposes of teaching, learning and research.

## 1. Background

In August 2010, the Council on Library and Information Resources (CLIR) and the Library of Congress (LoC) issued a report titled *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age* (2010). This report argues that our sound heritage continues to deteriorate on legacy formats making digitization of the utmost importance, but that preservation and access cannot be solved through digitization alone. As Mark Greene and Dennis Meissner remark in their 2005 article “More Product, Less Process,” “processing is not keeping up with acquisitions, and has not been for decades, resulting in massive backlogs of inaccessible collections at repositories across the country” (p. 208). The same is true of unprocessed and therefore inaccessible archives of spoken word sound collections that hold important cultural artifacts.

The 2010 report suggests that if scholars and students do not use sound archives, our cultural heritage institutions will be less inclined to preserve them. Consequently, archives and libraries must collaborate with patrons and scholars to understand how recordings are and might be used. If librarians and archivists need to know what scholars and students want to do with sound artifacts in order to make these collections more accessible, then humanities scholars, arguably, also need to know what kinds of analysis are possible in an age of large, freely available collections and advanced computational analysis.

Software development for accessing and analyzing audio in general is underdeveloped, however. Observing the general dead air in this audio access and preservation soundscape, CLIR’s *Survey of the State of Audio Collections in Academic Libraries* (Smith, et. al, 2004) and CLIR’s report with the LoC, *National Recording Preservation Plan* (Nelson-Strauss, et. al, 2012), cite copyright legislation reform, organizational initiatives for shared preservation networks, and improvements in the processes of discovery and cataloging as the areas where research and development are most needed for increasing access. They call for “new technologies for audio capture and automatic metadata extraction” (Smith, et. al, 2004, 11) with a “focus on developing, testing, and enhancing science-based approaches to all areas that affect audio preservation” (Nelson-Strauss, et. al, 2012, 15) to help relieve these dark backlogs of undescribed, even though digitized, audio collections.

Currently, there are a few free open-source content management systems that enhance access to audio and video in well-designed environments. For example, both the Avalon Media System at Indiana and Northwestern and the Oral History Metadata Synchronization project (OHMS) out of the University of Kentucky are open source systems specifically designed for managing large collections of digital audio and video files. Avalon and OHMS enable users to curate, distribute and provide online access to their collections for purposes of teaching, learning and research.

Other systems allow users to segment or to create clips and playlists for audio organization such as the Stories Matter Project at Concordia. Finally, proprietary tools such as Kairos and Glifos are content management tools that help automate production and cataloguing on the backend while providing for rich media delivery over diverse data transport platforms and presentation devices. With these tools, varied materials that all relate to a single event can be linked and presented together, but these tools are designed to leverage human-generated texts and metadata. Systems that are dependent on such content do not scale well in a culture that is constantly producing large digital sound collections without transcripts or reliable metadata.

An opportunity for changing human-intensive practices of annotating and indexing sound seems present in the fact that computer performance—in terms of speed, storage capacity, and advancements in machine learning—has increased. The very popular Digging into Data Challenge, which is supported by funding agencies representing Canada, the Netherlands, the United Kingdom, and the United States is a testament to the wide array of perspectives and methodologies digital projects can encompass. The first (2009) and second (2011) rounds of awards include projects that are using machine learning and visualization to provide new methods of discovery. Some analyze image files (“Digging into Image Data to Answer Authorship Related Questions”) and the word in text files (“Mapping the Republic of Letters” and “Using Zotero and TAPoR on the Old Bailey Proceedings: Data Mining with Criminal Intent”). Other Digging into Data projects provide new methods for discovery with audio files by analyzing large amounts of music information such as the “Structural Analysis of Large Amounts of Music” (SALAMI) and “the Electronic Locator of Vertical Interval Successions (ELVIS)” projects or—large scale data analysis of natural language usage (the “Mining a Year of Speech” and the “Harvesting Speech Datasets for Linguistic Research on the Web” projects). None of these projects, however, take on the perspective of scholars and cultural heritage institutions focused on folk, literary, or tribal spoken word collections. The third round of this funding, announced in January 2014, does not include any sound projects at all.

At the same time, the work required for such inquiry is related to the work that scholars have been doing for decades on features of music. The Automated Learning Group (ALG) of the Illinois Informatics Institute (I3) (where ARLO has been developed), for example has years of experience in high-performance machine learning and audio analysis. J. Stephen Downie (PI of SALAMI) and Michael Welge developed a system for comparing different music information retrieval (MIR) systems. Downie and collaborators also developed NEMA (Networked Environment for Music Analysis), a multinational, multidisciplinary cyber-infrastructure project for music information processing that builds upon music information retrieval research. NEMA can be used for genre and mood classification as well as composer identification; for similarity retrieval where similarity is measured on prosodic features of pitch, tempo, and accent or the key or tone of music; and

structural segmentation evaluation that identifies the key structural sections in music such as a change in verse, movement, or the addition of a chorus. Scholarship produced by Downie and collaborators as a result of their interventions in music information retrieval (Downie, 2003, 2008; Downie, et al., 2010) was particularly helpful in guiding our infrastructure development in the HiPSTAS project.

## **2. Project Activities**

The HiPSTAS institute had two primary learning outcomes: (A) participants would produce new scholarship using audio collections with advanced technologies such as classification, clustering, and visualizations; and (B) participants would engage in the scholarly work of digital infrastructure development by contributing to recommendations for the implementation of a suite of tools for collecting institutions interested in supporting advanced digital scholarship in sound.

NEH support for HiPSTAS was used primarily to set up ARLO for testing with spoken word audio collections, to conduct two HiPSTAS meetings (including travel for participants and speakers), to support consulting and software maintenance over the course of the year for the participants, and to support project management and administrative activities.

### **Pre-Institute Activities**

In the first nine months of funding for HiPSTAS, we created an online presence for reporting and general information:

- about the project at <http://blogs.ischool.utexas.edu/hipstas/about/>
- Call for Participants at <http://blogs.ischool.utexas.edu/hipstas/institute/cfp/> [also attached in Appendix F.]
- participant project pages at <https://sites.google.com/site/nehhipstas/documentation>
- Notes from both virtual and in-person meetings:  
<https://sites.google.com/site/nehhipstas/archive>
- a discussion forum at <https://groups.google.com/forum/?fromgroups=#!forum/hipstas> .

Clement and the UT GRA developed, advertised, and implemented an online application process for participants (shown in the link above). Our advisory board subsequently vetted and chose 20 participants, including humanities junior and senior faculty and advanced graduate students as well as librarians and archivists from across the U.S. who are interested in developing and using new technologies to access and analyze spoken word recordings within audio collections.

Our twenty participants came from a wide range of types of institutions and professional backgrounds: • Jeffrey Boruszak: PhD Candidate, Department of English, University of Texas-Austin • Kim Christen: Associate Professor and Director of Digital Projects, Washington State University • Hartwell Francis: Assistant Professor, Anthropology and Sociology, Western Carolina University • Michael Hennessey: Editor, PennSound and Jacket2 • Jennifer Himmelreich: MLIS Student, San Jose State University • Kira B. Homo: Electronic Records Archivist, University of Oregon • Justin Kovar: Digitization Project Archivist, The Dolph Briscoe Center For American History • Michael J. Kramer: Lecturer, Northwestern University • Bert Lyons: Digital Archivist, Library of Congress • Stephen McLaughlin: Director, PennSound Radio; Podcasts Editor, Jacket2.org • Ben Miller: Assistant Professor, Department of English, Georgia State University • Virginia Millington: Recording and Archive Manager, StoryCorps • Michael Nardone: PhD Candidate, Concordia University • Linda Newman: Digital Projects Coordinator, Langsam Library, University of Cincinnati • Juliana Nykolaiszyn: Assistant Professor; Oral History Librarian, Oklahoma State University • Amber Paraniek: Librarian, Library of Congress • Gena Peone: Assistant Cultural Collections Manager, Cultural Preservation Department, Spokane Tribe of Indians • Eric Rettberg: Postdoctoral Preceptor, University of Virginia • Elizabeth Russey Roke: Digital Archivist, Emory University • Kristen Suagee-Beauduy: Graduate Assistant, Department of English, Western Carolina University • Kenneth Sherwood: Associate Professor of English; Co-Director of Center for Digital Humanities and Culture, Indiana University of PA • Dustin Tahmahkera: Assistant Professor, Department of Communication Studies, Humanities Division, Southwestern University.

The participants' previous experience with audio analysis had been limited to the textual content of the audio files or, in some cases, waveform representations. The applications they were using (such as Audacity and iTunes) provided them with very limited access to the collections' sound features, features that Charles Bernstein and others have identified as significant for analysis. Bernstein (2011) claims that waveforms can only identify part of what makes poetry audio files interesting. He explains that "[t]here are four features or vocal gestures, that are available on tape but not page that are of special significance for poetry;" these include "the cluster of rhythm and tempo (including word duration), the cluster of pitch and intonation (including amplitude), timbre, and accent" (p. 144). The features that signify meaning in sound are diverse and a waveform can only visualize part of the first cluster (tempo) and part of the second cluster (amplitude). As such, though they were accustomed to working daily with large spoken text audio collections, prior to the HiPSTAS Institute, participants had never had access to the types of sound features machine-learning tools rely on for analyzing audio.

During the Institute, participants were introduced to ARLO (Adaptive Recognition with Layered Optimization), a machine learning application for analyzing large sound collections that was

originally developed to classify and analyze bird calls by extracting audio features and displaying the audio data as a spectral graph (Tcheng, et al., 2009 and Enstrom, et al., 2008). ARLO was then extended with National Science Foundation (NSF) funding to help another set of scholars classify pollen grains using image features instead of audio features (Tcheng, et al., 2008). ARLO extracts basic prosodic features such as pitch, rhythm and timbre for visualizations and spectral matching. These features are then used for classification and clustering. HiPSTAS shows that ARLO, which has been proven effective in the sciences, can also be used as a supercomputing resource for analyzing patterns across spoken word collections that are of interest to humanists.

Clement was approved through IRB at UT Austin to conduct usability studies with the HiPSTAS participants, with whom she conducted and recorded hour-long interviews before the Institute asking questions about their practices with access and scholarship with spoken word collections (see Appendix C for IRB application and interview script). Through subsequent interviews, observations and surveys, Clement gathered feedback throughout the HiPSTAS Institute in order to help inform development and evaluation.

#### *ARLO Software Development*

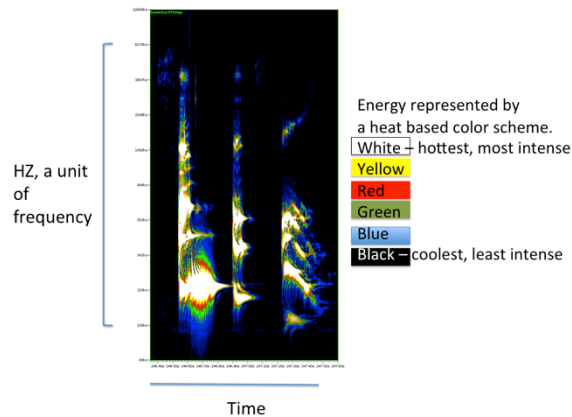
Participants primarily used the ARLO (Adaptive Recognition with Layered Optimization) software to analyze sound files since the I3 team saw its development as more advantageous to the project than developing NEMA, which was no longer under development at the start of this project.

ARLO, which was originally called NESTER, was developed with UIUC seed funding for avian ecologist David Enstrom (2008) to begin exploring the use of machine learning for data analysis in the fields of animal behavior and ecology. Specifically, by extracting basic prosodic features such as pitch, rhythm and timbre for classification, clustering, and visualizations, Enstrom used ARLO to identify and catalog all syllables (phonemes) produced by a species of songbird, the Northern Cardinal. With over 2400 hours of recordings, Enstrom used the pattern recognition and machine learning functions of ARLO to automatically produce bird vocalization transcripts and analyze vocal patterns in the audio data streams. As such, ARLO allowed Enstrom to process these streams and test hypotheses regarding song production and culture in birds that have heretofore been intractable.

ARLO analyzes event wave forms (raw audio samples) by extracting time and frequency information in the form of a spectrogram (see Figure 1). The spectrogram is computed using band pass filters linked with energy detectors. For example, in the spectrogram shown in Figure 1, the color of each pixel represents the numerical value of energy (which represents the sum of potential and kinetic energy) of a particular frequency at a point in time. Using a heat based color scheme,



the lowest values are black (cool), blue, green, red, yellow, and the points with the highest or most intense energy values are white. ARLO spectrograms contain similar information as FFT based spectrograms and are flexible in application because the frequencies and damping factors can be optimized for a given problem.



**Figure 1: A spectrogram produced by ARLO of Gertrude Stein saying “some such thing” from a reading of her novel *The Making of Americans*.**

The machine-learning algorithm ARLO uses for classification to find events in audio is called “instance based learning” (IBL). In IBL, the machine memorizes a number of classified training examples and matches them against new unseen examples to predict events. In ARLO, examples are audio events defined by a start and end time such as a two-second clip. ARLO finds matches by taking each known classified example and “sliding” it across new audio files looking for good matches. The number of match positions considered per second is adjustable and is set to the spectra sample rate. The degree of match or “match strength” is measured by the correlation between spectrograms, which can range between -1.0 and +1.0. In addition to simple spectra matching, a user can isolate pitch and volume traces, compute correlations on them, and weight the different feature types when computing the overall match strength. This allows the user to preferentially weight spectral information that might correspond to such aspects as pitch or rhythm.

For clustering, ARLO uses a single-threaded algorithm, ARLO selects a randomly chosen subset of examples then divides these examples into piles or clusters of equal or unequal size. ARLO evaluates the quality of every cluster by first computing the cluster centroid as an average of all the spectra and then averaging the distance of the examples in each cluster to the computed centroid. Finally, ARLO iterates over the following steps a given number of times:

1. randomly select two clusters and an example from each cluster to evaluate;

2. compute the cohesiveness of these two clusters based on exchanging the selected examples;
3. if the cohesiveness of the clustering improves, then exchange these two examples

Developing ARLO as a web-based application that can be leveraged by a wide community of users has been an essential goal of the HiPSTAS project. Before the HiPSTAS project, the ARLO front-end (Python/Django web interface) and back-end (Java processes and MySQL database) ran on a single dedicated server hosted by the National Center for Supercomputing Applications (NCSA) at UIUC. As a community tool, we have implemented ARLO on Stampede, a National Science Foundation funded petascale High Performance Computing (HPC) system at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. All users and projects are on the same machine and each user can create projects to keep their data private or share projects with each other. Having a single installation minimized maintenance and deployment times during development, but the Stampede server can handle the workload from user interactions on the front-end for numerous users at once.

Because HiPSTAS participants were able to use the ARLO implementation on Stampede, we were able to gather user and technical requirements for further developing ARLO for super computing tasks of interest to humanists working with spoken word sound collections. When describing the inherent difficulties in developing music information retrieval systems, Downie (2008) and Downie, et al. (2010) identify similar issues we faced when developing ARLO as a community tool. Downie identifies ten major research issues that must be addressed when developing MIRs including determining effective procedures and evaluation techniques for (1) indexing; (2) retrieval queries; (3) user interface design for access and analysis; (4) audio compression for efficient processing; (5) audio feature detection that yields productive analyses; (6) machine learning algorithms; (7) classification techniques; (8) security measures for sensitive materials; (9) accessibility procedures for a range of user communities; and (10) sufficient computing and storage infrastructure development for data-intensive techniques (Downie, 2008; Downie, et al., 2010). As described in the next section, our concern thus far in HiPSTAS has been focused on gathering user requirements for determining audio features that yield productive queries and infrastructure requirements that support appropriate computing and storage needs.

This development work for HiPSTAS included limited interface development with ARLO for non-birding humanities users, such as the ability to analyze longer files, adding short keys for play, stop, fast-forward, etc. This minor interface development allowed humanities users to test the machine learning system and perform exploratory discovery (clustering) and automated classification (prediction or supervised learning) processes as well as visualizations. Further, infrastructure development that allowed multiple users to use multiple collections and create

separate tags or annotation sets and share them. Finally, the project consultant, Tony Borries, Clement, Tcheng, and the UT GRA spent significant time on creating documentation for ARLO including:

- Documentation specifically designed for humanists interested in analyzing spoken text audio collections including step-by-step instructions as well as instructional videos: <https://sites.google.com/site/nehhipstas/documentation>.
- Documentation for developers: [http://wiki.arloproject.com/Main\\_Page](http://wiki.arloproject.com/Main_Page)

The original implementation of ARLO for modeling ran in parallel on systems at the National Center for Supercomputing Applications (NCSA). As part of HiPSTAS, the ARLO backend (written in Java) was developed to make calls to the Texas Advanced Computing Center's Stampede system.

For the HiPSTAS Institute, infrastructure development work was necessary in order to implement the ARLO backend on the Stampede system. This development included developing task scheduling technology, separating out and distributing Java processes, and setting up community user accounts for job batching and task management. Transitioning ARLO, which was implemented on a single processor, to achieve parallelism on Stampede necessitated developing a task manager to create sub problems and a task handler to do the work and return the results.

For the Institutes, we ingested 27,000 files from PennSound and 150 hours of folklore from the Dolph Briscoe Center for American History and set up user accounts for the HiPSTAS participants. Hooked up to Stampede, ARLO could accomplish computational tasks that required more processing power such as finding patterns across 27,000 PennSound audio files. Reading this amount of data, without precomputing or indexing would have taken days on a regular system, which would have precluded our ability to implement a sandbox for humanities scholars at the Austin-based Institutes.

### **First HiPSTAS Institute, May 29 – June 1, 2013**

The first four-day meeting of HiPSTAS held met at the University of Texas, May 29 – June 1, 2013, participants were introduced to essential issues that archivists, librarians, humanities scholars, and computer scientists and technologists face in understanding the nature of digital sound scholarship and the possibilities of building an infrastructure for enabling such scholarship. The Co-PI's developed a workshop to introduce participants to advanced computational analytics such as clustering, classification, and visualizations with ARLO (<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/first-meeting-may-29-june-1-2013/>). See **Appendix A** for first meeting agenda.

In the first year, Clement was also a significant spokesperson for the project, giving multiple talks about the project at various venues including scholarly conferences, universities and more popular venues:

*Invited Talks and Panels (first year, by Tanya Clement)*

- Plenary, “‘This is Just to Say’: Changing the Nature of Poetry Performance Studies and Learning to Listen with Machines,” University of Cincinnati (November 2013).
- Plenary, “Project-based DH in the Undergraduate Curriculum: a History, a Few Principles, and Some Suggestions” Arthur Vining Davis Digital Humanities Summer Faculty Workshop, Northwestern University (August 2013).
- Plenary, “Poetry in Motion: Using High Performance Sound Technologies for Access and Scholarship with Spoken-Word Collections” Poetry vs. Philosophy: Life, Artifact & Theory Symposium, Texas A&M University (April 2013).
- Invited talk, “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” University of Victoria, British Columbia, Canada (March 2013).
- Invited talk, “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” Approaching The Poetry Series: Using Literary Recordings as Scholars and Digital Designers, Concordia University, Montreal, Quebec, Canada (March 2013).
- Keynote: “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” Nebraska Forum on Digital Humanities “Hacking at Books”, Center for Digital Research in the Humanities. University of Nebraska – Lincoln (February 2013).

*Refereed Conference Panels and Presentations (first year, by Tanya Clement)*

- “What We Talk About When We Talk About Sound.” Digital Frontiers Conference, University of North Texas (September 2013).
- “Sound Seeings or High Performance Sound Technologies for Access and Scholarship” SXSWInteractive, Austin, TX (March 2013).

**Virtual Meetings, Interim year, June 2013 to May 2014**

Public virtual meetings (notes: <https://sites.google.com/site/nehhipstas/archive/virtual-meetings>) continued monthly over the course of the year with project team members and participants calling in.

Over the course of the year, the HiPSTAS teamwork yielded three significant results for the computational analysis of spoken word collections of keen interest to the humanities: (1) an assessment of user requirements; (2) an assessment of technological infrastructure needed to support a community tool; (3) preliminary experiments using these advanced resources that show the efficacy, both in terms of user needs and computational resources required, of using machine learning tools to improve discovery with unprocessed audio collections [these experiments and results appear in Clement et. al, 2014.]

## **Second HiPSTAS Institute, May 29 – June 1, 2014.**

In the second Institute, project team members discussed the evolution of the project and participants presented their work in a public event on the first day. On the second day, participants met to discuss recommendations for the continued development of High Performance Sound Technologies for Access and Scholarship. See **Appendix B** for second meeting agenda.

In the second year, Clement continued to speak for the project, giving multiple talks about the project at various venues including scholarly conferences, universities and more popular venues:

### *Invited Talks and Panels (second year, by Tanya Clement)*

- Invited talk, “The Aporia of Sound: Digital Humanities Futures in Poetry Performance Studies” Cornell University. The Conversations in Digital Humanities speaker series. Cornell University, Ithaca, NY (November 2015).
- Invited talk, “Part of Algorithmic Audition: A Roundtable on DH and Sound. The Center for Digital Humanities @ Princeton, Princeton University, Princeton, New Jersey (November 2015).
- Invited talk, “High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Text Mining In The Humanities: Collaboration across Sectors, University of Ottawa, May 31, 2015.
- Invited talk, “High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Digital Asset Symposium (DAS) (May 2015).
- Invited talk, “HiPSTAS: High Performance Sound Technologies for Access and Scholarship,” Cultural Analytics Conference, University of Chicago (May 2015).
- Invited talk, “Towards a Rationale of Audio Text: Audio Exploration and High Performance Technologies,” Penn Digital Humanities Forum, University of Pennsylvania (April 2015).
- Keynote, “The Shunting Yard: Developing Infrastructures for Meaning Making in Information and Sound Studies.” Texas Digital Humanities Conference, University of Houston (April 2014).
- Invited talk, “What We Talk about When We Talk about Sound: Introducing High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Sound + Conference, Center for Literary and Comparative and Studies of the English Department, University of Maryland, College Park (March 2014).
- Invited talk, “HiPSTAS, What?: Information Retrieval, Machine Learning, and Visualizations with Sound,” CUNY Digital Humanities Initiative, New York (March 2014).
- Invited talk, “Introducing High Performance Sound Technologies for Access and Scholarship,” Columbia University, New York (March 2014).

### *Refereed Conference Panels and Presentations (second year, by Tanya Clement)*

- “Machinic Ballads: Alan Lomax’s Global Jukebox and the Categorization of Sound Culture” Society for Ethnomusicology Annual Conference, Austin, TX, (December 2015).
- “Reconsidering the Hermeneutics of Listening with High Performance Sound Technologies for Access and Scholarship (HiPSTAS). Inertia: A Conference on Sound, Media, and the Digital Humanities, University of California, Los Angeles (May 2015).
- “Do you Hear What I Hear? Introducing High Performance Sound Technologies for Access and Scholarship.” Society of American Archivists Annual Meeting, Washington DC (August

- 2014).
- “Developing for Distant Listening: Developing Computational Tools for Sound Analysis By Framing User Requirements within Critical Theories for Sound Studies” Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “<audio>Digital Humanities</audio>: The Intersections of Sound and Method.” Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “The Hermeneutics of Distant Listening to Spoken Word Texts with High Performance Sound Technologies for Access and Scholarship” AV in DH, 2014 Workshop, Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “Using Sound Technologies in the Study and Documentation of Spoken Word Recordings.” Native American and Indigenous Studies (NAISA) Annual Meeting, Austin, TX (May 2014).

*Publications by and about HiPSTAS:*

- Clement, T., McLaughlin, S., Tchong, D., Auvil, L., Borries, T. “Measured Applause: Toward a Cultural Analysis of Audio Collections” *Cultural Analytics* (under review).
- Clement, T. “A Rationale of Audio Text.” *digital humanities quarterly* (under review).
- Clement, T. “The Ear and the Shunting Yard: Meaning Making as Resonance in Early Information Theory.” *Information & Culture* 49.4 (2014): 401-426.
- Clement, T. “Word. Spoken. Articulating the Voice as Descriptive Metadata for High Performance Sound Technologies for Access and Scholarship (HiPSTAS).” In *Provoke*. Darren Mueller, Mary Caton Lingold and Whitney Anne Trettien (eds.) Durham, NC: Duke University Press (Forthcoming).
- Clement, T. “Introducing High Performance Sound Technologies for Access and Scholarship.” *The International Association of Sound and Audiovisual Archives Journal* (September 2013) 41: 21-28.
- Clement, T. “When Texts of Study are Audio Files: Digital Tools for Sound Studies in DH” In Susan Schreibman, Ray Siemens and John Unsworth (eds.), *A New Companion to Digital Humanities* (Blackwell Companions to Literature and Culture), 2016: 348-357.
- Clement, T. “Machinic Ballads: Alan Lomax’s Global Jukebox and the Categorization of Sound Culture.” *Sounding Out! Blog*. 14 May 2015.
- Clement, T., Tchong, D., Auvil, L., and Borries, T. “High Performance Sound Technologies for Access and Scholarship (HiPSTAS) in the Digital Humanities” Proceedings of the 77th Annual ASIST Conference, Seattle, WA, 31 October – 5 November.
- Clement, T. and Roy, L., “HiPSTAS: An Institute Advancing Tools for Analyzing Digital Audio Collections,” *American Indian Library Association Newsletter* 36 (2013): 8-15.
- Filreis, A. “Anti-ordination in the visualization of the poem's sound” *Jacket2*. 27 February 2014. <http://jacket2.org/commentary/anti-finality-visualization-poems-sound>
- Francis\*, H., Clement, T., Peone, G., Carpenter, B., Suagee-Beauduy, K. “Accessing Sound at Libraries, Archives, and Museums” *Indigenous Ownership & Libraries, Archives, and Museums* (forthcoming).
- MacArthur, M. “Monotony, the Churches of Poetry Reading, and Sound Studies.” *PMLA*. Forthcoming 2016.
- Mustazza, C. ” “The noise is the content: Toward computationally determining the provenance of poetry recordings" using ARLO. *Jacket2*. 10 Jan. 2015. <https://jacket2.org/commentary/noise-content-toward-computationally-determining-provenance-poetry-recordings>
- Perez-Hernandez, D. “Scholars Collaborate to Make Sound Recordings More Accessible” *The Chronicle of Higher Education*. 26 March 2014. <http://chronicle.com/blogs/wiredcampus/scholars-collaborate-to-make-sound-recordings-more-accessible/51215>
- Rettberg, E. “Hearing the Audience”. *Jacket2*. 26 March 2015.

- <http://jacket2.org/commentary/hearing-audience>
- Sherwood, K. “Distanced sounding: ARLO as a tool for the analysis and visualization of versioning phenomena within poetry audio” *Jacket2*. 2 March 2015.  
<https://jacket2.org/commentary/distanced-sounding-arlo-tool-analysis-and-visualization-versioning-phenomena-within-poetr>

#### 4. Accomplishments

We planned several important results from the HiPSTAS Institute. The curriculum of this Institute is premised on the idea that building scholarly infrastructure in the digital humanities is the work of scholars, librarians and archivists, and computer scientists together.

Our resulting scholarship (listed above) reflects deep collaborations across scholars, computer scientists and cultural heritage professionals and a knowledge of digital sound preservation and computational analysis of sound. Further, the participating collections have increased their ability to allow users to perform new kinds of scholarship with the data sets we created during the grant period. Finally, the recommendations we have created for the development of advanced computational tools for digital scholarly inquiry in sound reflect the needs and concerns of both the stewards of sound collections and the scholars who use them. The publications listed above describe three use cases: (1) Poetry (Clement, “Word. Spoken.”; Clement, “When Texts of Study are Audio Files” and Clement, “Towards a Rationale of Audiotext”); (2) Folklore (Clement, “Machinic Ballads”); and (3) Recordings in Indigenous Communities (Francis\*, H., Clement, T., Peone, G., Carpenter, B., Suagee-Beauduy, K. “Accessing Sound at Libraries, Archives, and Museums”). The next NEH grant, described below, funds use case work we are currently doing with Archivists.

#### 5. Audiences

Specifically, the HiPSTAS Institute represented a wide range of professional communities that are necessarily impacted by increased access and scholarship with sound. Our 20 participants included:

- 9 librarians and archivists
- 8 humanities scholars
- 3 advanced graduate students in humanities and information science

These students, scholars, and practitioners represented interests in audio collections from diverse communities across the United States including Native American tribal communities and Civil Rights collections from the American South. The projects to which we had access during the Institute (namely, 30,000 audio files from PennSound and 57 feet of tapes (reels and audiocassettes) from the UT folklore collection at the Dolph the Briscoe Center for American History at UT Austin each represented a wide range of voices from across the United States. Further, participants came from communities that represent an even wider range of voices and communities. Collections associated with our participants include but are not limited to:

- Field recordings (200,000 recordings) American Folklife Center, Library of Congress
- 30,000 hours, Oral histories, Storycorps
- Speeches in the Southern Christian Leadership Conference recordings, Emory University
- 700 recordings in the Elliston Poetry Collection at the University of Cincinnati

- 36 interviews in the Dust, Drought and Dreams Gone Dry: Oklahoma Women and the Dust Bowl (WDB) oral history project out of the Oklahoma State Libraries

New audiences include many partners from a diverse range of disciplines (biology, humanities, linguistics, and libraries and archives); our Advisory Board which includes representatives who use and represent important AV archives and libraries such as the Archive of Indigenous Languages of Latin America (AILLA), the Macaulay Library of Natural Sounds, and the WGBH Media Library and Archives; and representatives from national aggregators such as DPLA, HTDL and the HTRC; as well as metadata and AV experts involved in the Europeana Sounds project, AVPreserve, CLIR, Pop Up Archive, and the UT iSchool.

## 6. Evaluation

The project was continually evaluated by project team members, project participants, and third party reviewers of presentations at conferences and scholarly publication venues as evidenced by the list of publications and presentations in Section 1: Project Activities. Evaluations were given to all the project participants at the second meeting Institute meeting in Austin, Texas in May, 2014. These evaluations were done in groups. The questions we asked included the following:

### *Group Evaluations*

1. In general, what are the ideal tools that would be most valuable to your community?
2. In particular, how can a tool like ARLO accommodate current workflow practices?
3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?
4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?
5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?
6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?
7. What are future grant or collaboration opportunities the group might engage to support this work?
8. What are future pedagogical or publishing opportunities the group might engage to support this work?
9. What aspect of the process of developing a tool like ARLO is/was most frustrating? How do we ameliorate
10. What are more questions we should be asking? (answer them, please)

### *Evaluation Results*

The results of the evaluation were that participants were enthusiastic about the continued development of ARLO and the use of such tools in their own research, in classrooms, with community members, and with collections management. The participants gave us copious response on what features were needed for further developing ARLO. These are included in Appendix D. More qualitative results for each of the questions above appear in Appendix E with four groups represented in these accounts: Archivists; Poetry Librarians and Scholars; Native American Librarians and Scholars; and Sound Researchers.



The project team's assessment of the program was very positive. It would have ultimately been useful to have more development money for ARLO, since it is not stable enough to share with scholars beyond the institute, but this has been the focus of subsequent grants that have taken advantage of both the content of these initial conversations as well as the relationships that were created.

#### *Public Response*

The public response was also very positive as evidenced by the Chronicle of Higher Education publication (cited above), as well as well-attended presentations at more popular venues such as SXSW and DAS. Further, because of her project leadership, Clement was asked to become a Research Associate for the Radio Preservation Task Force of the Library of Congress.

### **7. Continuation of the Project**

The success of the project is evidenced by the second grant we were awarded by the National Endowment for the Humanities Preservation and Access grant for "HiPSTAS Research and Development with Repositories" (HRDR). In this HRDR phase, we plan to leverage the conversations and collaborations we established as part of HiPSTAS to develop the ARLO software as a more generally accessible and usable tool for the wider humanities community at both small and large institutions and to teach these communities how to use it. To this end, the HRDR phase will include three primary products: (1) a release of ARLO (Automated Recognition with Layered Optimization) that leverages machine learning and visualizations to augment the creation of descriptive metadata for use with a variety of repositories (such as a MySQL database, Fedora, or CONTENTdm); (2) a workshop curriculum and documentation for wider dissemination and training with the software; and (4) a white paper that details best practices for automatically generating descriptive metadata for spoken word digital audio collections in the humanities.

HiPSTAS participant Marit MacArthur has received an ACLS digital innovation fellowship to develop the ARLO interface for humanists interested in pitch tracking. She is actively pursuing other granting possibilities.

The use of ARLO and development under Clement's direction plays a small part in the tools being developed in the recently IMLS-funded project involving WGBH and the Pop Up Archive for "Improving Access to Time-Based Media through Crowdsourcing and Machine Learning" (on which Clement is an Advisory Board member).

Finally, we have just applied for a third major grant through the Institute of Museum and Library Services (IMLS) for funds to implement and test implementing ARLO at the Texas Advanced Computing Center at the University of Texas. Like many tools developed for research, however, ARLO 1.0 lacks essential aspects for broader implementation such as user-tested interfaces and workflows that reflect the storage capacity and processing power needed to efficiently meet long-term demands for real users in a wide range of settings who want to access AV materials. This partnership includes many partners from a diverse range of disciplines (biology, humanities, linguistics, and libraries and archives); our Advisory Board which includes representatives who use and represent important AV archives and libraries such as the Archive of Indigenous Languages of Latin America (AILLA), the Macaulay Library of Natural Sounds, and the WGBH Media Library and Archives; and representatives from national aggregators such as DPLA, HTDL and the HTRC; as well as metadata and AV experts involved in the Europeana Sounds project, AVPreserve, CLIR, Pop Up Archive, and the UT iSchool.

To better understand and document the socio-technical needs associated with trying new, largescale, and machine-automated processes, our proposal seeks funding for completing the following deliverables:

- 1) A suite of tested interfaces for ARLO 2.0, including documentation, tutorials, and sample AV files.
- 2) A suite of tested machine learning algorithms for searching and identifying significant patterns in AV collections, including documentation, tutorials, and sample data sets.
- 3) A suite of API-driven executable code bases for implementing ARLO 2.0 on personal computers, local or cloud servers, and on supercomputer clusters, including documentation.
- 4) Reports and recommendations for implementing ARLO 2.0 at different scales for researchers from the varied fields represented by our use cases and for archives and libraries who seek to make their AV collections discoverable as part of the National Digital Platform. The reports will include sample workflows for mapping data generated by ARLO into DPLA Metadata Application Profile (Map) 4.0, the MARC21 format (which is used by HTDL), and PBCore, which is used by WGBH and the Pop Up Archive as well as recommendations for the creation of feature sets (as a means for offering information about collections while also preserving copyright and privacy restrictions)

## 8. Long Term Impact

Evidence of projects that have been inspired by the HiPSTAS Institute appear in the Native American, Archives, Poetry, and Sound Studies communities appear above in the list of publications. Currently, ARLO is not usable beyond a small set of people, which negatively impacts our ability to create a wider community of users and the impact this work might have in the classroom and in scholarship. At the same time, as the publications show, there is great interest. Anecdotal evidence also suggests that HiPSTAS participants have been applying for grant funding for their own sound projects.

## 9. Grant Products

We anticipated three primary outcomes of the HiPSTAS Institute, closely tied to evaluation, as having the most impact on advancing computational analysis in sound scholarship. These results have been achieved:

- a) **Publicly available ongoing evaluation on the process of scholarship and technical developments** including the website at the UT iSchool (<http://blogs.ischool.utexas.edu/hipstas/> and related sites linked there) that we have maintained as a public source for information about the project and a venue for disseminating final reports. Much of the dialog before the first Institute, in the interim year, and after the close of the Institute took place in project work spaces we created in free, open-source platforms on Google (<https://sites.google.com/site/nehhipstas/project-pages>). Beyond the monthly virtual meetings and the big mid-January 2014 meeting in Google Hangout, we have used Google Sites for an ongoing master document in which the project team has included documentation for Humanists seeking to use the ARLO software (<https://sites.google.com/site/nehhipstas/documentation>) as well as for developers ([http://wiki.arloproject.com/Main\\_Page](http://wiki.arloproject.com/Main_Page)).
- b) **Curriculum and Scholarship:** The curriculum, including the ARLO labs, and the outcome of both meetings of the HiPSTAS Institute have been made openly available as part of the iSchool web site

(<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/first-meeting-may-29-june-1-2013/>) and (<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/second-meeting-may-27-28-2014/>). Subsequent to the final meeting, participants were invited to contribute scholarship to a special series of *Jacket2* magazine (a preeminent and open source venue for creative and scholarly, digital work) on experimental digital analyses of poetry audio, titled “Clippings” (these are listed in “Publications” above). The visibility of the curricular materials, and of participants’ samples and documentation via the web site and open source publication venues have made the results of this Institute accessible to a wider audience beyond those able to participate directly.

- c) **Final white paper and recommendations:** The final white paper, written by Clement, Tchong and Auvil, reflects the monthly status reports and the developing infrastructure in ARLO. This report includes recommendations for implementing an open-source, freely available suite of tools for supporting scholarship on audio files. The purpose of disseminating these recommendations through the UT iSchool is to offer best practices for cultural heritage institutes that are new to making their sound files available via Web services frameworks and to provide the final recommendations for developing and implementing a more robust technical infrastructure based on feedback about ARLO collected during the HipSTAS Institute. The report is being disseminated at <http://blogs.ischool.utexas.edu/hipstas/2016/01/25/hipstas-neh-institute-final-white-paper/>
- d) **ARLO code** (<https://bitbucket.org/arloproject/> )

## 10. Appendices

Appendix A: First Meeting Agenda, May 29 – June 1, 2013; University of Texas at Austin

Appendix B: Second Meeting Agenda, May 27 – 28, 2014; UT Austin iSchool,

Appendix C: IRB Application with Interview Protocol

Appendix D: ARLO Requirements for Humanists

Appendix E: Group Evaluations of ARLO

Appendix F: Call for Participants

Other systems allow users to segment or to create clips and playlists for audio organization such as the Stories Matter Project at Concordia. Finally, proprietary tools such as Kairos and Glifos are content management tools that help automate production and cataloguing on the backend while providing for rich media delivery over diverse data transport platforms and presentation devices. With these tools, varied materials that all relate to a single event can be linked and presented together, but these tools are designed to leverage human-generated texts and metadata. Systems that are dependent on such content do not scale well in a culture that is constantly producing large digital sound collections without transcripts or reliable metadata.

An opportunity for changing human-intensive practices of annotating and indexing sound seems present in the fact that computer performance—in terms of speed, storage capacity, and advancements in machine learning—has increased. The very popular Digging into Data Challenge, which is supported by funding agencies representing Canada, the Netherlands, the United Kingdom, and the United States is a testament to the wide array of perspectives and methodologies digital projects can encompass. The first (2009) and second (2011) rounds of awards include projects that are using machine learning and visualization to provide new methods of discovery. Some analyze image files (“Digging into Image Data to Answer Authorship Related Questions”) and the word in text files (“Mapping the Republic of Letters” and “Using Zotero and TAPoR on the Old Bailey Proceedings: Data Mining with Criminal Intent”). Other Digging into Data projects provide new methods for discovery with audio files by analyzing large amounts of music information such as the “Structural Analysis of Large Amounts of Music” (SALAMI) and “the Electronic Locator of Vertical Interval Successions (ELVIS)” projects or—large scale data analysis of natural language usage (the “Mining a Year of Speech” and the “Harvesting Speech Datasets for Linguistic Research on the Web” projects). None of these projects, however, take on the perspective of scholars and cultural heritage institutions focused on folk, literary, or tribal spoken word collections. The third round of this funding, announced in January 2014, does not include any sound projects at all.

At the same time, the work required for such inquiry is related to the work that scholars have been doing for decades on features of music. The Automated Learning Group (ALG) of the Illinois Informatics Institute (I3) (where ARLO has been developed), for example has years of experience in high-performance machine learning and audio analysis. J. Stephen Downie (PI of SALAMI) and Michael Welge developed a system for comparing different music information retrieval (MIR) systems. Downie and collaborators also developed NEMA (Networked Environment for Music Analysis), a multinational, multidisciplinary cyber-infrastructure project for music information processing that builds upon music information retrieval research. NEMA can be used for genre and mood classification as well as composer identification; for similarity retrieval where similarity is measured on prosodic features of pitch, tempo, and accent or the key or tone of music; and

structural segmentation evaluation that identifies the key structural sections in music such as a change in verse, movement, or the addition of a chorus. Scholarship produced by Downie and collaborators as a result of their interventions in music information retrieval (Downie, 2003, 2008; Downie, et al., 2010) was particularly helpful in guiding our infrastructure development in the HiPSTAS project.

## **2. Project Activities**

The HiPSTAS institute had two primary learning outcomes: (A) participants would produce new scholarship using audio collections with advanced technologies such as classification, clustering, and visualizations; and (B) participants would engage in the scholarly work of digital infrastructure development by contributing to recommendations for the implementation of a suite of tools for collecting institutions interested in supporting advanced digital scholarship in sound.

NEH support for HiPSTAS was used primarily to set up ARLO for testing with spoken word audio collections, to conduct two HiPSTAS meetings (including travel for participants and speakers), to support consulting and software maintenance over the course of the year for the participants, and to support project management and administrative activities.

### **Pre-Institute Activities**

In the first nine months of funding for HiPSTAS, we created an online presence for reporting and general information:

- about the project at <http://blogs.ischool.utexas.edu/hipstas/about/>
- Call for Participants at <http://blogs.ischool.utexas.edu/hipstas/institute/cfp/> [also attached in Appendix F.]
- participant project pages at <https://sites.google.com/site/nehhipstas/documentation>
- Notes from both virtual and in-person meetings:  
<https://sites.google.com/site/nehhipstas/archive>
- a discussion forum at <https://groups.google.com/forum/?fromgroups=#!forum/hipstas> .

Clement and the UT GRA developed, advertised, and implemented an online application process for participants (shown in the link above). Our advisory board subsequently vetted and chose 20 participants, including humanities junior and senior faculty and advanced graduate students as well as librarians and archivists from across the U.S. who are interested in developing and using new technologies to access and analyze spoken word recordings within audio collections.

Our twenty participants came from a wide range of types of institutions and professional backgrounds: • Jeffrey Boruszak: PhD Candidate, Department of English, University of Texas-Austin • Kim Christen: Associate Professor and Director of Digital Projects, Washington State University • Hartwell Francis: Assistant Professor, Anthropology and Sociology, Western Carolina University • Michael Hennessey: Editor, PennSound and Jacket2 • Jennifer Himmelreich: MLIS Student, San Jose State University • Kira B. Homo: Electronic Records Archivist, University of Oregon • Justin Kovar: Digitization Project Archivist, The Dolph Briscoe Center For American History • Michael J. Kramer: Lecturer, Northwestern University • Bert Lyons: Digital Archivist, Library of Congress • Stephen McLaughlin: Director, PennSound Radio; Podcasts Editor, Jacket2.org • Ben Miller: Assistant Professor, Department of English, Georgia State University • Virginia Millington: Recording and Archive Manager, StoryCorps • Michael Nardone: PhD Candidate, Concordia University • Linda Newman: Digital Projects Coordinator, Langsam Library, University of Cincinnati • Juliana Nykolaiszyn: Assistant Professor; Oral History Librarian, Oklahoma State University • Amber Paraniek: Librarian, Library of Congress • Gena Peone: Assistant Cultural Collections Manager, Cultural Preservation Department, Spokane Tribe of Indians • Eric Rettberg: Postdoctoral Preceptor, University of Virginia • Elizabeth Russey Roke: Digital Archivist, Emory University • Kristen Suagee-Beauduy: Graduate Assistant, Department of English, Western Carolina University • Kenneth Sherwood: Associate Professor of English; Co-Director of Center for Digital Humanities and Culture, Indiana University of PA • Dustin Tahmahkera: Assistant Professor, Department of Communication Studies, Humanities Division, Southwestern University.

The participants' previous experience with audio analysis had been limited to the textual content of the audio files or, in some cases, waveform representations. The applications they were using (such as Audacity and iTunes) provided them with very limited access to the collections' sound features, features that Charles Bernstein and others have identified as significant for analysis. Bernstein (2011) claims that waveforms can only identify part of what makes poetry audio files interesting. He explains that "[t]here are four features or vocal gestures, that are available on tape but not page that are of special significance for poetry;" these include "the cluster of rhythm and tempo (including word duration), the cluster of pitch and intonation (including amplitude), timbre, and accent" (p. 144). The features that signify meaning in sound are diverse and a waveform can only visualize part of the first cluster (tempo) and part of the second cluster (amplitude). As such, though they were accustomed to working daily with large spoken text audio collections, prior to the HiPSTAS Institute, participants had never had access to the types of sound features machine-learning tools rely on for analyzing audio.

During the Institute, participants were introduced to ARLO (Adaptive Recognition with Layered Optimization), a machine learning application for analyzing large sound collections that was

originally developed to classify and analyze bird calls by extracting audio features and displaying the audio data as a spectral graph (Tcheng, et al., 2009 and Enstrom, et al., 2008). ARLO was then extended with National Science Foundation (NSF) funding to help another set of scholars classify pollen grains using image features instead of audio features (Tcheng, et al., 2008). ARLO extracts basic prosodic features such as pitch, rhythm and timbre for visualizations and spectral matching. These features are then used for classification and clustering. HiPSTAS shows that ARLO, which has been proven effective in the sciences, can also be used as a supercomputing resource for analyzing patterns across spoken word collections that are of interest to humanists.

Clement was approved through IRB at UT Austin to conduct usability studies with the HiPSTAS participants, with whom she conducted and recorded hour-long interviews before the Institute asking questions about their practices with access and scholarship with spoken word collections (see Appendix C for IRB application and interview script). Through subsequent interviews, observations and surveys, Clement gathered feedback throughout the HiPSTAS Institute in order to help inform development and evaluation.

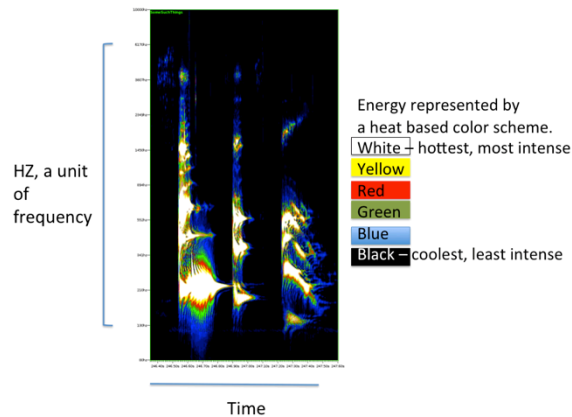
#### *ARLO Software Development*

Participants primarily used the ARLO (Adaptive Recognition with Layered Optimization) software to analyze sound files since the I3 team saw its development as more advantageous to the project than developing NEMA, which was no longer under development at the start of this project.

ARLO, which was originally called NESTER, was developed with UIUC seed funding for avian ecologist David Enstrom (2008) to begin exploring the use of machine learning for data analysis in the fields of animal behavior and ecology. Specifically, by extracting basic prosodic features such as pitch, rhythm and timbre for classification, clustering, and visualizations, Enstrom used ARLO to identify and catalog all syllables (phonemes) produced by a species of songbird, the Northern Cardinal. With over 2400 hours of recordings, Enstrom used the pattern recognition and machine learning functions of ARLO to automatically produce bird vocalization transcripts and analyze vocal patterns in the audio data streams. As such, ARLO allowed Enstrom to process these streams and test hypotheses regarding song production and culture in birds that have heretofore been intractable.

ARLO analyzes event wave forms (raw audio samples) by extracting time and frequency information in the form of a spectrogram (see Figure 1). The spectrogram is computed using band pass filters linked with energy detectors. For example, in the spectrogram shown in Figure 1, the color of each pixel represents the numerical value of energy (which represents the sum of potential and kinetic energy) of a particular frequency at a point in time. Using a heat based color scheme,

the lowest values are black (cool), blue, green, red, yellow, and the points with the highest or most intense energy values are white. ARLO spectrograms contain similar information as FFT based spectrograms and are flexible in application because the frequencies and damping factors can be optimized for a given problem.



**Figure 1: A spectrogram produced by ARLO of Gertrude Stein saying “some such thing” from a reading of her novel *The Making of Americans*.**

The machine-learning algorithm ARLO uses for classification to find events in audio is called “instance based learning” (IBL). In IBL, the machine memorizes a number of classified training examples and matches them against new unseen examples to predict events. In ARLO, examples are audio events defined by a start and end time such as a two-second clip. ARLO finds matches by taking each known classified example and “sliding” it across new audio files looking for good matches. The number of match positions considered per second is adjustable and is set to the spectra sample rate. The degree of match or “match strength” is measured by the correlation between spectrograms, which can range between -1.0 and +1.0. In addition to simple spectra matching, a user can isolate pitch and volume traces, compute correlations on them, and weight the different feature types when computing the overall match strength. This allows the user to preferentially weight spectral information that might correspond to such aspects as pitch or rhythm.

For clustering, ARLO uses a single-threaded algorithm, ARLO selects a randomly chosen subset of examples then divides these examples into piles or clusters of equal or unequal size. ARLO evaluates the quality of every cluster by first computing the cluster centroid as an average of all the spectra and then averaging the distance of the examples in each cluster to the computed centroid. Finally, ARLO iterates over the following steps a given number of times:

1. randomly select two clusters and an example from each cluster to evaluate;



2. compute the cohesiveness of these two clusters based on exchanging the selected examples;
3. if the cohesiveness of the clustering improves, then exchange these two examples

Developing ARLO as a web-based application that can be leveraged by a wide community of users has been an essential goal of the HiPSTAS project. Before the HiPSTAS project, the ARLO front-end (Python/Django web interface) and back-end (Java processes and MySQL database) ran on a single dedicated server hosted by the National Center for Supercomputing Applications (NCSA) at UIUC. As a community tool, we have implemented ARLO on Stampede, a National Science Foundation funded petascale High Performance Computing (HPC) system at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. All users and projects are on the same machine and each user can create projects to keep their data private or share projects with each other. Having a single installation minimized maintenance and deployment times during development, but the Stampede server can handle the workload from user interactions on the front-end for numerous users at once.

Because HiPSTAS participants were able to use the ARLO implementation on Stampede, we were able to gather user and technical requirements for further developing ARLO for super computing tasks of interest to humanists working with spoken word sound collections. When describing the inherent difficulties in developing music information retrieval systems, Downie (2008) and Downie, et al. (2010) identify similar issues we faced when developing ARLO as a community tool. Downie identifies ten major research issues that must be addressed when developing MIRs including determining effective procedures and evaluation techniques for (1) indexing; (2) retrieval queries; (3) user interface design for access and analysis; (4) audio compression for efficient processing; (5) audio feature detection that yields productive analyses; (6) machine learning algorithms; (7) classification techniques; (8) security measures for sensitive materials; (9) accessibility procedures for a range of user communities; and (10) sufficient computing and storage infrastructure development for data-intensive techniques (Downie, 2008; Downie, et al., 2010). As described in the next section, our concern thus far in HiPSTAS has been focused on gathering user requirements for determining audio features that yield productive queries and infrastructure requirements that support appropriate computing and storage needs.

This development work for HiPSTAS included limited interface development with ARLO for non-birding humanities users, such as the ability to analyze longer files, adding short keys for play, stop, fast-forward, etc. This minor interface development allowed humanities users to test the machine learning system and perform exploratory discovery (clustering) and automated classification (prediction or supervised learning) processes as well as visualizations. Further, infrastructure development that allowed multiple users to use multiple collections and create

separate tags or annotation sets and share them. Finally, the project consultant, Tony Borries, Clement, Tcheng, and the UT GRA spent significant time on creating documentation for ARLO including:

- Documentation specifically designed for humanists interested in analyzing spoken text audio collections including step-by-step instructions as well as instructional videos: <https://sites.google.com/site/nehhipstas/documentation>.
- Documentation for developers: [http://wiki.arloproject.com/Main\\_Page](http://wiki.arloproject.com/Main_Page)

The original implementation of ARLO for modeling ran in parallel on systems at the National Center for Supercomputing Applications (NCSA). As part of HiPSTAS, the ARLO backend (written in Java) was developed to make calls to the Texas Advanced Computing Center's Stampede system.

For the HiPSTAS Institute, infrastructure development work was necessary in order to implement the ARLO backend on the Stampede system. This development included developing task scheduling technology, separating out and distributing Java processes, and setting up community user accounts for job batching and task management. Transitioning ARLO, which was implemented on a single processor, to achieve parallelism on Stampede necessitated developing a task manager to create sub problems and a task handler to do the work and return the results.

For the Institutes, we ingested 27,000 files from PennSound and 150 hours of folklore from the Dolph Briscoe Center for American History and set up user accounts for the HiPSTAS participants. Hooked up to Stampede, ARLO could accomplish computational tasks that required more processing power such as finding patterns across 27,000 PennSound audio files. Reading this amount of data, without precomputing or indexing would have taken days on a regular system, which would have precluded our ability to implement a sandbox for humanities scholars at the Austin-based Institutes.

### **First HiPSTAS Institute, May 29 – June 1, 2013**

The first four-day meeting of HiPSTAS held met at the University of Texas, May 29 – June 1, 2013, participants were introduced to essential issues that archivists, librarians, humanities scholars, and computer scientists and technologists face in understanding the nature of digital sound scholarship and the possibilities of building an infrastructure for enabling such scholarship. The Co-PI's developed a workshop to introduce participants to advanced computational analytics such as clustering, classification, and visualizations with ARLO (<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/first-meeting-may-29-june-1-2013/>). See **Appendix A** for first meeting agenda.

In the first year, Clement was also a significant spokesperson for the project, giving multiple talks about the project at various venues including scholarly conferences, universities and more popular venues:

*Invited Talks and Panels (first year, by Tanya Clement)*

- Plenary, “‘This is Just to Say’: Changing the Nature of Poetry Performance Studies and Learning to Listen with Machines,” University of Cincinnati (November 2013).
- Plenary, “Project-based DH in the Undergraduate Curriculum: a History, a Few Principles, and Some Suggestions” Arthur Vining Davis Digital Humanities Summer Faculty Workshop, Northwestern University (August 2013).
- Plenary, “Poetry in Motion: Using High Performance Sound Technologies for Access and Scholarship with Spoken-Word Collections” Poetry vs. Philosophy: Life, Artifact & Theory Symposium, Texas A&M University (April 2013).
- Invited talk, “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” University of Victoria, British Columbia, Canada (March 2013).
- Invited talk, “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” Approaching The Poetry Series: Using Literary Recordings as Scholars and Digital Designers, Concordia University, Montreal, Quebec, Canada (March 2013).
- Keynote: “Sound Seeings or High Performance Sound Technologies for Access and Scholarship.” Nebraska Forum on Digital Humanities “Hacking at Books”, Center for Digital Research in the Humanities. University of Nebraska – Lincoln (February 2013).

*Refereed Conference Panels and Presentations (first year, by Tanya Clement)*

- “What We Talk About When We Talk About Sound.” Digital Frontiers Conference, University of North Texas (September 2013).
- “Sound Seeings or High Performance Sound Technologies for Access and Scholarship” SXSWInteractive, Austin, TX (March 2013).

**Virtual Meetings, Interim year, June 2013 to May 2014**

Public virtual meetings (notes: <https://sites.google.com/site/nehhipstas/archive/virtual-meetings>) continued monthly over the course of the year with project team members and participants calling in.

Over the course of the year, the HiPSTAS teamwork yielded three significant results for the computational analysis of spoken word collections of keen interest to the humanities: (1) an assessment of user requirements; (2) an assessment of technological infrastructure needed to support a community tool; (3) preliminary experiments using these advanced resources that show the efficacy, both in terms of user needs and computational resources required, of using machine learning tools to improve discovery with unprocessed audio collections [these experiments and results appear in Clement et. al, 2014.]

## **Second HiPSTAS Institute, May 29 – June 1, 2014.**

In the second Institute, project team members discussed the evolution of the project and participants presented their work in a public event on the first day. On the second day, participants met to discuss recommendations for the continued development of High Performance Sound Technologies for Access and Scholarship. See **Appendix B** for second meeting agenda.

In the second year, Clement continued to speak for the project, giving multiple talks about the project at various venues including scholarly conferences, universities and more popular venues:

### *Invited Talks and Panels (second year, by Tanya Clement)*

- Invited talk, “The Aporia of Sound: Digital Humanities Futures in Poetry Performance Studies” Cornell University. The Conversations in Digital Humanities speaker series. Cornell University, Ithaca, NY (November 2015).
- Invited talk, “Part of Algorithmic Audition: A Roundtable on DH and Sound. The Center for Digital Humanities @ Princeton, Princeton University, Princeton, New Jersey (November 2015).
- Invited talk, “High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Text Mining In The Humanities: Collaboration across Sectors, University of Ottawa, May 31, 2015.
- Invited talk, “High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Digital Asset Symposium (DAS) (May 2015).
- Invited talk, “HiPSTAS: High Performance Sound Technologies for Access and Scholarship,” Cultural Analytics Conference, University of Chicago (May 2015).
- Invited talk, “Towards a Rationale of Audio Text: Audio Exploration and High Performance Technologies,” Penn Digital Humanities Forum, University of Pennsylvania (April 2015).
- Keynote, “The Shunting Yard: Developing Infrastructures for Meaning Making in Information and Sound Studies.” Texas Digital Humanities Conference, University of Houston (April 2014).
- Invited talk, “What We Talk about When We Talk about Sound: Introducing High Performance Sound Technologies for Access and Scholarship (HiPSTAS),” Sound + Conference, Center for Literary and Comparative and Studies of the English Department, University of Maryland, College Park (March 2014).
- Invited talk, “HiPSTAS, What?: Information Retrieval, Machine Learning, and Visualizations with Sound,” CUNY Digital Humanities Initiative, New York (March 2014).
- Invited talk, “Introducing High Performance Sound Technologies for Access and Scholarship,” Columbia University, New York (March 2014).

### *Refereed Conference Panels and Presentations (second year, by Tanya Clement)*

- “Machinic Ballads: Alan Lomax’s Global Jukebox and the Categorization of Sound Culture” Society for Ethnomusicology Annual Conference, Austin, TX, (December 2015).
- “Reconsidering the Hermeneutics of Listening with High Performance Sound Technologies for Access and Scholarship (HiPSTAS). Inertia: A Conference on Sound, Media, and the Digital Humanities, University of California, Los Angeles (May 2015).
- “Do you Hear What I Hear? Introducing High Performance Sound Technologies for Access and Scholarship.” Society of American Archivists Annual Meeting, Washington DC (August

- 2014).
- “Developing for Distant Listening: Developing Computational Tools for Sound Analysis By Framing User Requirements within Critical Theories for Sound Studies” Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “<audio>Digital Humanities</audio>: The Intersections of Sound and Method.” Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “The Hermeneutics of Distant Listening to Spoken Word Texts with High Performance Sound Technologies for Access and Scholarship” AV in DH, 2014 Workshop, Digital Humanities Conference, Lausanne, Switzerland (July 2014).
  - “Using Sound Technologies in the Study and Documentation of Spoken Word Recordings.” Native American and Indigenous Studies (NAISA) Annual Meeting, Austin, TX (May 2014).

*Publications by and about HiPSTAS:*

- Clement, T., McLaughlin, S., Tchong, D., Auvil, L., Borries, T. “Measured Applause: Toward a Cultural Analysis of Audio Collections” *Cultural Analytics* (under review).
- Clement, T. “A Rationale of Audio Text.” *digital humanities quarterly* (under review).
- Clement, T. “The Ear and the Shunting Yard: Meaning Making as Resonance in Early Information Theory.” *Information & Culture* 49.4 (2014): 401-426.
- Clement, T. “Word. Spoken. Articulating the Voice as Descriptive Metadata for High Performance Sound Technologies for Access and Scholarship (HiPSTAS).” In *Provoke*. Darren Mueller, Mary Caton Lingold and Whitney Anne Trettien (eds.) Durham, NC: Duke University Press (Forthcoming).
- Clement, T. “Introducing High Performance Sound Technologies for Access and Scholarship.” *The International Association of Sound and Audiovisual Archives Journal* (September 2013) 41: 21-28.
- Clement, T. “When Texts of Study are Audio Files: Digital Tools for Sound Studies in DH” In Susan Schreibman, Ray Siemens and John Unsworth (eds.), *A New Companion to Digital Humanities* (Blackwell Companions to Literature and Culture), 2016: 348-357.
- Clement, T. “Machinic Ballads: Alan Lomax’s Global Jukebox and the Categorization of Sound Culture.” *Sounding Out! Blog*. 14 May 2015.
- Clement, T., Tchong, D., Auvil, L., and Borries, T. “High Performance Sound Technologies for Access and Scholarship (HiPSTAS) in the Digital Humanities” Proceedings of the 77th Annual ASIST Conference, Seattle, WA, 31 October – 5 November.
- Clement, T. and Roy, L., “HiPSTAS: An Institute Advancing Tools for Analyzing Digital Audio Collections,” *American Indian Library Association Newsletter* 36 (2013): 8-15.
- Filreis, A. “Anti-ordination in the visualization of the poem's sound” *Jacket2*. 27 February 2014. <http://jacket2.org/commentary/anti-finality-visualization-poems-sound>
- Francis\*, H., Clement, T., Peone, G., Carpenter, B., Suagee-Beauduy, K. “Accessing Sound at Libraries, Archives, and Museums” *Indigenous Ownership & Libraries, Archives, and Museums* (forthcoming).
- MacArthur, M. “Monotony, the Churches of Poetry Reading, and Sound Studies.” *PMLA*. Forthcoming 2016.
- Mustazza, C. ” “The noise is the content: Toward computationally determining the provenance of poetry recordings" using ARLO. *Jacket2*. 10 Jan. 2015. <https://jacket2.org/commentary/noise-content-toward-computationally-determining-provenance-poetry-recordings>
- Perez-Hernandez, D. “Scholars Collaborate to Make Sound Recordings More Accessible” *The Chronicle of Higher Education*. 26 March 2014. <http://chronicle.com/blogs/wiredcampus/scholars-collaborate-to-make-sound-recordings-more-accessible/51215>
- Rettberg, E. “Hearing the Audience”. *Jacket2*. 26 March 2015.

- <http://jacket2.org/commentary/hearing-audience>
- Sherwood, K. “Distanced sounding: ARLO as a tool for the analysis and visualization of versioning phenomena within poetry audio” *Jacket2*. 2 March 2015.  
<https://jacket2.org/commentary/distanced-sounding-arlo-tool-analysis-and-visualization-versioning-phenomena-within-poetr>

#### 4. Accomplishments

We planned several important results from the HiPSTAS Institute. The curriculum of this Institute is premised on the idea that building scholarly infrastructure in the digital humanities is the work of scholars, librarians and archivists, and computer scientists together.

Our resulting scholarship (listed above) reflects deep collaborations across scholars, computer scientists and cultural heritage professionals and a knowledge of digital sound preservation and computational analysis of sound. Further, the participating collections have increased their ability to allow users to perform new kinds of scholarship with the data sets we created during the grant period. Finally, the recommendations we have created for the development of advanced computational tools for digital scholarly inquiry in sound reflect the needs and concerns of both the stewards of sound collections and the scholars who use them. The publications listed above describe three use cases: (1) Poetry (Clement, “Word. Spoken.”; Clement, “When Texts of Study are Audio Files” and Clement, “Towards a Rationale of Audiotext”); (2) Folklore (Clement, “Machinic Ballads”); and (3) Recordings in Indigenous Communities (Francis\*, H., Clement, T., Peone, G., Carpenter, B., Suagee-Beauduy, K. “Accessing Sound at Libraries, Archives, and Museums”). The next NEH grant, described below, funds use case work we are currently doing with Archivists.

#### 5. Audiences

Specifically, the HiPSTAS Institute represented a wide range of professional communities that are necessarily impacted by increased access and scholarship with sound. Our 20 participants included:

- 9 librarians and archivists
- 8 humanities scholars
- 3 advanced graduate students in humanities and information science

These students, scholars, and practitioners represented interests in audio collections from diverse communities across the United States including Native American tribal communities and Civil Rights collections from the American South. The projects to which we had access during the Institute (namely, 30,000 audio files from PennSound and 57 feet of tapes (reels and audiocassettes) from the UT folklore collection at the Dolph the Briscoe Center for American History at UT Austin each represented a wide range of voices from across the United States. Further, participants came from communities that represent an even wider range of voices and communities. Collections associated with our participants include but are not limited to:

- Field recordings (200,000 recordings) American Folklife Center, Library of Congress
- 30,000 hours, Oral histories, Storycorps
- Speeches in the Southern Christian Leadership Conference recordings, Emory University
- 700 recordings in the Elliston Poetry Collection at the University of Cincinnati

- 36 interviews in the Dust, Drought and Dreams Gone Dry: Oklahoma Women and the Dust Bowl (WDB) oral history project out of the Oklahoma State Libraries

New audiences include many partners from a diverse range of disciplines (biology, humanities, linguistics, and libraries and archives); our Advisory Board which includes representatives who use and represent important AV archives and libraries such as the Archive of Indigenous Languages of Latin America (AILLA), the Macaulay Library of Natural Sounds, and the WGBH Media Library and Archives; and representatives from national aggregators such as DPLA, HTDL and the HTRC; as well as metadata and AV experts involved in the Europeana Sounds project, AVPreserve, CLIR, Pop Up Archive, and the UT iSchool.

## 6. Evaluation

The project was continually evaluated by project team members, project participants, and third party reviewers of presentations at conferences and scholarly publication venues as evidenced by the list of publications and presentations in Section 1: Project Activities. Evaluations were given to all the project participants at the second meeting Institute meeting in Austin, Texas in May, 2014. These evaluations were done in groups. The questions we asked included the following:

### *Group Evaluations*

1. In general, what are the ideal tools that would be most valuable to your community?
2. In particular, how can a tool like ARLO accommodate current workflow practices?
3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?
4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?
5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?
6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?
7. What are future grant or collaboration opportunities the group might engage to support this work?
8. What are future pedagogical or publishing opportunities the group might engage to support this work?
9. What aspect of the process of developing a tool like ARLO is/was most frustrating? How do we ameliorate
10. What are more questions we should be asking? (answer them, please)

### *Evaluation Results*

The results of the evaluation were that participants were enthusiastic about the continued development of ARLO and the use of such tools in their own research, in classrooms, with community members, and with collections management. The participants gave us copious response on what features were needed for further developing ARLO. These are included in Appendix D. More qualitative results for each of the questions above appear in Appendix E with four groups represented in these accounts: Archivists; Poetry Librarians and Scholars; Native American Librarians and Scholars; and Sound Researchers.

The project team's assessment of the program was very positive. It would have ultimately been useful to have more development money for ARLO, since it is not stable enough to share with scholars beyond the institute, but this has been the focus of subsequent grants that have taken advantage of both the content of these initial conversations as well as the relationships that were created.

#### *Public Response*

The public response was also very positive as evidenced by the Chronicle of Higher Education publication (cited above), as well as well-attended presentations at more popular venues such as SXSW and DAS. Further, because of her project leadership, Clement was asked to become a Research Associate for the Radio Preservation Task Force of the Library of Congress.

### **7. Continuation of the Project**

The success of the project is evidenced by the second grant we were awarded by the National Endowment for the Humanities Preservation and Access grant for "HiPSTAS Research and Development with Repositories" (HRDR). In this HRDR phase, we plan to leverage the conversations and collaborations we established as part of HiPSTAS to develop the ARLO software as a more generally accessible and usable tool for the wider humanities community at both small and large institutions and to teach these communities how to use it. To this end, the HRDR phase will include three primary products: (1) a release of ARLO (Automated Recognition with Layered Optimization) that leverages machine learning and visualizations to augment the creation of descriptive metadata for use with a variety of repositories (such as a MySQL database, Fedora, or CONTENTdm); (2) a workshop curriculum and documentation for wider dissemination and training with the software; and (4) a white paper that details best practices for automatically generating descriptive metadata for spoken word digital audio collections in the humanities.

HiPSTAS participant Marit MacArthur has received an ACLS digital innovation fellowship to develop the ARLO interface for humanists interested in pitch tracking. She is actively pursuing other granting possibilities.

The use of ARLO and development under Clement's direction plays a small part in the tools being developed in the recently IMLS-funded project involving WGBH and the Pop Up Archive for "Improving Access to Time-Based Media through Crowdsourcing and Machine Learning" (on which Clement is an Advisory Board member).

Finally, we have just applied for a third major grant through the Institute of Museum and Library Services (IMLS) for funds to implement and test implementing ARLO at the Texas Advanced Computing Center at the University of Texas. Like many tools developed for research, however, ARLO 1.0 lacks essential aspects for broader implementation such as user-tested interfaces and workflows that reflect the storage capacity and processing power needed to efficiently meet long-term demands for real users in a wide range of settings who want to access AV materials. This partnership includes many partners from a diverse range of disciplines (biology, humanities, linguistics, and libraries and archives); our Advisory Board which includes representatives who use and represent important AV archives and libraries such as the Archive of Indigenous Languages of Latin America (AILLA), the Macaulay Library of Natural Sounds, and the WGBH Media Library and Archives; and representatives from national aggregators such as DPLA, HTDL and the HTRC; as well as metadata and AV experts involved in the Europeana Sounds project, AVPreserve, CLIR, Pop Up Archive, and the UT iSchool.



To better understand and document the socio-technical needs associated with trying new, largescale, and machine-automated processes, our proposal seeks funding for completing the following deliverables:

- 1) A suite of tested interfaces for ARLO 2.0, including documentation, tutorials, and sample AV files.
- 2) A suite of tested machine learning algorithms for searching and identifying significant patterns in AV collections, including documentation, tutorials, and sample data sets.
- 3) A suite of API-driven executable code bases for implementing ARLO 2.0 on personal computers, local or cloud servers, and on supercomputer clusters, including documentation.
- 4) Reports and recommendations for implementing ARLO 2.0 at different scales for researchers from the varied fields represented by our use cases and for archives and libraries who seek to make their AV collections discoverable as part of the National Digital Platform. The reports will include sample workflows for mapping data generated by ARLO into DPLA Metadata Application Profile (Map) 4.0, the MARC21 format (which is used by HTDL), and PBCore, which is used by WGBH and the Pop Up Archive as well as recommendations for the creation of feature sets (as a means for offering information about collections while also preserving copyright and privacy restrictions)

## 8. Long Term Impact

Evidence of projects that have been inspired by the HiPSTAS Institute appear in the Native American, Archives, Poetry, and Sound Studies communities appear above in the list of publications. Currently, ARLO is not usable beyond a small set of people, which negatively impacts our ability to create a wider community of users and the impact this work might have in the classroom and in scholarship. At the same time, as the publications show, there is great interest. Anecdotal evidence also suggests that HiPSTAS participants have been applying for grant funding for their own sound projects.

## 9. Grant Products

We anticipated three primary outcomes of the HiPSTAS Institute, closely tied to evaluation, as having the most impact on advancing computational analysis in sound scholarship. These results have been achieved:

- a) **Publicly available ongoing evaluation on the process of scholarship and technical developments** including the website at the UT iSchool (<http://blogs.ischool.utexas.edu/hipstas/> and related sites linked there) that we have maintained as a public source for information about the project and a venue for disseminating final reports. Much of the dialog before the first Institute, in the interim year, and after the close of the Institute took place in project work spaces we created in free, open-source platforms on Google (<https://sites.google.com/site/nehhipstas/project-pages>). Beyond the monthly virtual meetings and the big mid-January 2014 meeting in Google Hangout, we have used Google Sites for an ongoing master document in which the project team has included documentation for Humanists seeking to use the ARLO software (<https://sites.google.com/site/nehhipstas/documentation>) as well as for developers ([http://wiki.arloproject.com/Main\\_Page](http://wiki.arloproject.com/Main_Page)).
- b) **Curriculum and Scholarship:** The curriculum, including the ARLO labs, and the outcome of both meetings of the HiPSTAS Institute have been made openly available as part of the iSchool web site

(<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/first-meeting-may-29-june-1-2013/>) and (<http://blogs.ischool.utexas.edu/hipstas/institute/meetings/second-meeting-may-27-28-2014/>). Subsequent to the final meeting, participants were invited to contribute scholarship to a special series of *Jacket2* magazine (a preeminent and open source venue for creative and scholarly, digital work) on experimental digital analyses of poetry audio, titled “Clippings” (these are listed in “Publications” above). The visibility of the curricular materials, and of participants’ samples and documentation via the web site and open source publication venues have made the results of this Institute accessible to a wider audience beyond those able to participate directly.

- c) **Final white paper and recommendations:** The final white paper, written by Clement, Tchong and Auvil, reflects the monthly status reports and the developing infrastructure in ARLO. This report includes recommendations for implementing an open-source, freely available suite of tools for supporting scholarship on audio files. The purpose of disseminating these recommendations through the UT iSchool is to offer best practices for cultural heritage institutes that are new to making their sound files available via Web services frameworks and to provide the final recommendations for developing and implementing a more robust technical infrastructure based on feedback about ARLO collected during the HipSTAS Institute. The report is being disseminated at <http://blogs.ischool.utexas.edu/hipstas/2016/01/25/hipstas-neh-institute-final-white-paper/>
- d) **ARLO code** (<https://bitbucket.org/arloproject/> )

## 10. Appendices

Appendix A: First Meeting Agenda, May 29 – June 1, 2013; University of Texas at Austin

Appendix B: Second Meeting Agenda, May 27 – 28, 2014; UT Austin iSchool,

Appendix C: IRB Application with Interview Protocol

Appendix D: ARLO Requirements for Humanists

Appendix E: Group Evaluations of ARLO

Appendix F: Call for Participants

**Appendix A: First Meeting: May 29 – June 1, 2013**  
University of Texas at Austin

Time

8:00 am -  
9:00 am

**Continental Breakfast**

*Location: Harry Ransom Center*

9:00 am -  
9:30 am

**Introductions, Overview of the HiPSTAS Institute** [*Tanya Clement and Loretta Auvil*]

**The State of Sound and Cultural Preservation** [*Sarah Cunningham and Lorieene Roy* ] Cunningham will introduce the field of sound preservation, including essential past and perceived future issues, and will discuss updates in the field since the publication of the National Recordings Preservation Board's project *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age* (August 2010). She will introduce how sound preservation is taught at the UT iSchool as a result of this publication, using LBJ Recordings as an example of collaborations in sound between scholars, sound archivists, and students. She will discuss her current work with the IMLS Oral History in the Digital Age Board and her chapter on audio preservation for Oral Historians. Roy will discuss recent work in cultural heritage initiatives and professional LIS (Library and Information Science) organizations concerning TCE (traditional cultural expressions) and the impact these conversations have on incorporating sound recordings from tribal communities into HiPSTAS. These organizations include American Indian Library Association (AILA), the Association of Tribal Libraries, Archives and Museums (ATALM), and the Tribal College Librarians Professional Development Institute among others. Roy will discuss the potential of negotiating access through the promise of *Mukurtu*, and contributions to Native language revitalization. Christen will introduce participants to the *Mukurtu* system. *Mukurtu* (<http://www.mukurtu.org/>) is a free and open source community content management system that provides international standards-based tools adaptable to the local cultural protocols and intellectual property systems of Indigenous communities, libraries, archives, and museums. Is a flexible archival tool that allows users to protect, preserve and share digital cultural heritage, and the representative will discuss a range of possible ways in which a HiPSTAS system could be developed to work with the *Mukurtu* system.

9:30 am -  
10:45 am

10:45 am -  
11:00 am

**Break**

11:00 am -

**The SALAMI Project and the State of Structural Analysis of Music** [*J. Stephen*

12:30 pm	<p><i>Downie</i>] Downie will discuss the current state of structural analysis of music (formal analysis), which is one of the most fundamental analyses performed by music researchers who seek to understand the overall view of a piece. Any course of formal analysis is often a core course in undergraduate music curricula. Formal analysis is useful in classifying different genres of music and it can be used to compare different styles of composition within a composer's works or between composers. It can also be used to understand historical influences over time and location. Downie will discuss the SALAMI Project (including the development of NEMA), the goal of which is to develop new text mining methods that are consistent with the manual processes that experts currently used to analyze music. Downie will discuss key outcomes from this study, including a longitudinal study of manual discovery and synthesis behaviors of a diverse network of faculty, policy makers, and students, advances in natural language processing methods that automatically identify concepts and relationships, detect entailment and paraphrasing, and generate multi-document summaries, a collection of gold standards that reflect diverse and realistic information needs that will drive further research in natural language processing. Downie will show examples of the analysis of large sets of music and new discoveries made with these questions.</p>
12:30 – 2 pm	<p><b>Box Lunches.</b> <i>HRC Tour #1, 1:30 to 2pm</i></p>
2 – 3:15 pm	<p><b>Introduction to HiPSTAS Participating Collections and Scholarly Perspectives on Sound Studies</b> [ <i>Al Filreis, Quinn Stewart, John Wheat</i>] Representatives from each of the participating collections will introduce these collections. Stewart will introduce LBJ recordings made available through Glifos. Filreis will discuss the poetry collections at PennSound. Gunn will introduce the folklore collections at the Briscoe Center.</p>
3:15 – 3:45 pm	<p><b>Break</b> <i>HRC Tour #2</i></p>
3:45 – 5 pm	<p><b>Introduction to Scholarly and Cultural Perspectives on Sound Studies</b> [<i>Steve Evans and Timothy Powell</i>] Evans will trace the emergence of the “phonotextuality” as an area of inquiry and analysis within the field of literary hermeneutics. He will discuss his own work on poetry audio files, which dates back to archival experiences with analog formats (mostly reel-to-reel and cassette) in the 1980s, and talk about the increased interest in recorded poetry among scholars (and poets) in the era of freely-accessible large-scale digital file serving platforms such as PennSound, Ubuweb, the Naropa Poetics Audio Archive, and many others. He will introduce the work of other scholars and poets interested in sound studies. Finally, he will address some of the challenges, and opportunities, involved in the attempt to adapt advanced computational tools to the purposes of humanistic inquiry in the field of poetry and poetics. Powell will introduce the storytelling audio collections of the Native American Projects at the American Philosophical Society. Powell will discuss the particular interests that Ojibwe people or scholars may have in analyzing sound files within “Gibagadinamaagoom: An Ojibwe Digital Archive” including examples of the shift</p>

from English to *Ojibwemowin* ('the Ojibwe language') at culturally significant moments. This shift is the work of the Ojibwe *oshkabewis* ("one empowered to translate between the spiritual and mundane worlds") and is of great interest both to elders who seek to educate youth in the ways of the language and for scholars and sound preservationists interested in analyzing the collections. In addition, Powell will discuss possible analysis that scholars interested in Native American Projects may have in doing analysis across collections of different tribes.

## Day 2: May 30, 2013

8:15 – 9 am

### **Continental Breakfast**

*Location: Texas Advanced Computing Center Visualization Lab [breakfast, morning session, and lunch]*

9 am – 10:45  
am

### **Introduction to Participant Projects: 3-Minute Lightning Rounds**

**Sound Visualizations and Visualizations Lab at TACC (Texas Advanced Computing Center)** [*Tanya Clement, David Tcheng, and Rob Turknett*] Rob Turknett will start out with a brief introduction to humanities visualization projects at TACC. Clement will introduce ProseVis, a tool developed in collaboration with the Auvil and Tcheng which considers machine learning and visualization with features of sound derived from text. Clement and Tcheng will give a brief history of ARLO as it corresponds to sound analysis and visualization technological history and computational visualizations of sound. Participants will be introduced to audio spectrogram visualizations of their audio files. Participants will see how different prosody features can be extracted from the spectrograms.

10:45 am -  
12:00 pm

12:00 pm -  
1:30 pm

Lunch

*Location: iSchool Lab*

**Using High Performance Sound Technologies, Introduction to Visualization and Discovery Processes in ARLO** [*David Tcheng*]. This lab will begin with a round table discussion of participant problems then a hands-on introduction to visualization in ARLO. Tcheng will demonstrate how to visualize frequency ranges including human hearing (20 – 20K) and microphone response (40 – 15K) as well as how to selecting

1:30 pm -  
4:30 pm

ranges appropriate to the task. He will demonstrate zooming in and out on time and frequency as well as the benefits of changing damping factors and changing gain. Tcheng will also introduce tagging and supervised tag discovery. Participants will conduct guided experiments in each task. For example, Tcheng will demonstrate similarity based search. Participants will identify (tag) a single segment of interest in their audio collection using ARLO's spectrogram visualizations as their guide. Given a single tagged example, participants will search their audio collections with ARLO finding most similar matches. Next we progress to predictive modeling (classification, supervised learning). Participants will be allowed to "tag" more examples in their collections creating multiple examples of each category of interest. The result of this tagging process will be a catalog of examples of each category. These examples will be transformed into a classification model using ARLO's predictive modeling capabilities. Finally, the classification model will be used to classify larger portions of entire collections to discover new patterns of interest.

### Day 3: May 31, 2013

8:30am-9:00  
am

#### **Continental Breakfast**

*Location: iSchool Lab*

#### **Using High Performance Sound Technologies, Introduction to Classification in ARLO**

[*David Tcheng*] Tcheng will describe unsupervised tag discovery with clustering in ARLO. He will demonstrate this process using examples from poetry, cardinals, and music. After a brief introduction to clustering, participants will perform clustering with existing tags from the previous days' lab. Then, participants will perform clustering with randomly selected windows.

9:00 am-  
12:00 pm

Location:

12:00 pm-  
1:30 pm

#### **Box Lunches**

#### **Developing Infrastructure with Use Cases in DH**

[*Tanya Clement and Loretta Auvil*] Clement and Auvil will introduce participants to the advantages and pitfalls of developing technical infrastructures with dispersed use cases. This time will include setting the groundwork for the online space in which the Institute will meet over the course of the year including introducing participants to the Google Sites space, how to post and edit on the wiki and establishing expectations for the monthly status reports. This space will provide a key component of the project

1:30 pm –  
2:00 pm

since it will mark the progress of the developing use cases and the developing augmentation of ARLO based on use case needs and become the basis of the final recommendations offered by the Institute for scholars, computer scientists, and librarians and archivists interested in participating in further development of the HiPSTAS infrastructure.

2:00 pm - 3:00 pm	Small group break-out discussions on defining use cases and re-articulating project goals with Co-PIs Clement and Auvil will also work with participants to re-articulate individual project goals and needed resources, and to create a project plan with proposed deadlines.
3:00 pm - 4:00 pm	Large group discussion on issues arising in small groups

#### Day 4: June 1, 2013

8:15 am - 9:00 am	<b>Continental Breakfast</b>
----------------------	------------------------------

*Location: iSchool Lab*

9:00 am - 12:00 pm	<b>Using High Performance Sound Technologies, Use Cases</b> [David Tcheng]. People will work with HiPSTAS team on their own use cases.
-----------------------	--

12:00 pm – 1pm	Box lunches and closing remarks
-------------------	---------------------------------

## **Appendix B:** Second Meeting, May 27 – 28, 2014

UT Austin iSchool,

Building UTA, Room 1.208

### **Day 1**

Time

8:30 am –

9:00 am

Continental Breakfast

9:00 am –

9:30 am

Dean Andrew Dillon welcome; Tanya Clement and Loretta Auvil: HiPSTAS  
Year One: What Did We Do?

#### **Michael Nardone** *Remarks on ARLO*

In these remarks, I hope to do a few things: to reflect on this past year of research activity and the confluence of this group of people who are all pursuing various aspects of recorded sound; to begin to frame the discussions that follow in terms of what kinds of research we might be able to produce through this confluence, individually and collaboratively; and, finally, to advocate for a kind of grey literature, a mode of technical reports that might facilitate a discussion of where we can go from here, what we can produce through our researches.

#### **Eric Rettberg** *Looking for Laughter in PennSound*

Given a tool that can search sound with other sounds, what kinds of sounds would we actually want to look for? While some have focused on the particularities of background noise or on the particularities of poets' speaking voices, I'll argue that the sound of laughter represents a compelling target. Laughter proves surprisingly central to the experience of reading or listening to poetry, and it's a form of response that has gone largely unrecorded in the age of the book. By exploring laughter with ARLO, we can better understand authors' irony and earnestness, find moments of distance between an author's intentions and an audience's response, and better understand the crucial role that laughter has played in twentieth- and twenty-first-century poetry.

#### **Ken Sherwood** *Distanced Sounding: ARLO Visualization and Poetry Audio Versioning*

One interest in spoken word poetry archives is the access to audio files documenting numerous performance instances of a given poem. How can audio visualization and tag discovery help to identify and represent degrees of variance between multiple instances? Which variant features (tempo, volume, pitch, rhythm ...) are most salient for the human listener, and which can be identified through visualization and machine learning?

9:30 am –

10:45 am

**Panel 1**

10:45 am –

11:00 am

Break



**Michael Kramer** *“Fishing Blues”: Using ARLO to Explore Musical Patterns of Community on Harry Smith’s Anthology of American Folk Music*

I have been using ARLO iteratively to search for unperceived connections and contrasts within a small data set: the famous Anthology of American Folk Music put together by Harry Smith in 1952. Consisting of 84 tracks, divided into three categories—Ballads, Social Music, and Songs—this collage of US “roots” music recordings was a kind of mystical remix of old commercial recordings of hillbilly sounds, race records, and other ethnic musics that existed on the periphery of the emerging twentieth-century American commercial recording industry.

**Juliana Nykolaiszyn** *In search of Oklahoma*

Pulling from oral histories recorded with women who survived the Dust Bowl in the 1930s, this presentation will feature preliminary tagging analysis exploring “Oklahoma” and other key words in this oral history collection.

**Elizabeth Roke** *ARLO for Archivists: The Potential of Machine Learning for Basic Metadata Generation*

When collections containing unlabeled, unidentified audio recordings arrive in the archives, archivists generally must listen to each recording to provide descriptive metadata for researchers including date, content, and type of recording. As a result, many thousands of hours of audio in archival repositories remain unidentified due to the scarcity of staff resources and are hidden from researchers. What if we could automate part of this process of description? This presentation explores the potential for using machine learning to help archivists expose these hidden audio collections and make them more widely available.

**Toneisha Taylor:** *Lomax, Cade and the Ethnography of Sound recordings: Investigation into the WPA Slave Narrative Collection*

Since the beginning of the WPA Folklore project and the private funding of the collection of oral histories and personal narratives from formally enslaved people in the 1920’s and 1930’s there has been some controversy about the collection and preservation of sound recordings. This project recognizes the controversy around the original recordings and asks critical questions about preservation, access and possibility of use(s) for the recordings. John Lomax’s work within the FWP at the helm of the Folklore and Folkways collection projects has meant the preservation of significant recordings about Black life in the Deep South during and after slavery. John B Cade, a contemporary of Lomax, and African American historian collected over 400 interviews in 13 states while at Prairie View State College (now Prairie View A&M University). Where Lomax’s collection was largely publicly funded, many parts of Cade’s collection were privately funded even though he was employed by a public college. The proposed project is an critical ethnography of the archive and sound recordings. The questions asked reflect critically upon both the Lomax and Cade collections. Questions of race, class, and gender are asked in both the collection and preservation of the collections. While both deal extensively with the collection of narratives from former slaves the bodies that interviewed and collected these narratives, the intent and training of those engaged in

11:00 am –  
12:30 pm  
**Panel 2**

collection, and the roles of those providing their narratives and personal histories matters in ways that impact current and past practices. The goal here is to use critical ethnography methods of collection and analysis to understand the ways we both silence and celebrate the voices within sound archives. ARLO could allow for the analysis of and comparisons between two very large corpora. It may be possible to identify locations of recordings, home locations of participants or duplications in narratives collected i.e. did the same person collect narratives for Cade and Lomax? Or, equally possible, did the same person give interviews to interviewers representing the projects directed by Lomax and Cade? What can we learn from a critical ethnography of sound when we compare these to large corpora.

12:30 pm –  
1:45 pm

Lunch: Austin's Pizza (Tocker Lounge, 1st Floor)

**Marit MacArthur** *Large-Scale Pitch Tracking in Poetry Recordings—How Can We Do This?*

Pattern and variation in pitch is arguably one of the most interesting aspects of performance styles. Compared to other aspects such as tempo and volume, however, tracking pitch presents unique obstacles for scholars, especially on a large scale, because of the particular qualities of the human voice. Errors in pitch-tracking are ubiquitous in audio analysis, and to make matters worse, most people are unreliable judges of pitch and change in pitch. Yet the common practice among linguistics who study acoustic phonetics is to correct for errors manually. In this talk, I will review the basics of acoustic phonetics and the problematic qualities of the human voice, and explore possibilities for using machine learning to correct errors in pitch-tracking and accurately identify patterns in pitch across a large number of audio recordings.

**Steve McLaughlin** *"Sound in Space: Crude Regression-Based Audio Classification"*

What can we learn about a recording without listening to it? Can we automatically sort audio based on genre, gender, and speech style? Is it possible for an algorithm to judge the quality of a poetry reading? This paper will examine distant reading strategies for audio in general and speech in particular, with an emphasis on high-dimensional regression models.

**Chris Mustazza:** *Forensic Audio Analysis to Determine the Provenance of Poetry Recordings*

Is it possible to determine the provenance of poetry recordings based on artifacts from the materiality of their recording histories? For example, if we know that one set of poetry recordings was created on a specific recording device, can we use that recording to identify other recordings in an archive that were recorded on the same device, thus helping to determine their provenance?

1:45 pm –  
3:00 pm  
**Panel 3**

3:00 pm –  
3:15 pm

Break

**Hartwell Francis** *Looking at Sound: Quick and Close*

Visualizations of the stream of sound facilitate rapid access and pattern identification. Digital graphic representations free users from the physiological constraints of sound stream processing while providing information about the sound stream that is mostly lost in written representations. Digital packaging provides rapid access to repeated patterns allowing users to isolate key sound stream segments. Through rapid access and pattern identification, visualizations of the Cherokee language stream of sound promote close listening to Cherokee language sound information. Close listening is a key component of language acquisition and language learning.

**Virginia Millington** *Accessing the StoryCorps Archive Using Digital Tools*

With over 30,000 hours of audio in the StoryCorps Archive, the collection represents an ideal data set in which to test a diverse array of digital tools of discovery. From ARLO to OHMS, StoryCorps has pursued ways in which to search, access, analyze and share the contents of this remarkable resource, including investigating methods of discovery that not only highlight the words spoken in any given interview, but also its content, tone, affect, emotion, and more. This talk will include specific examples of this work, as well as a discussion of what the future holds for the StoryCorps Archive.

**Tim Powell** *Tagging the Spirits: Using ARLO to Identify When the Spirit of the Drum Speaks to an Ojibwe Traditional Knowledge Keeper*

Tanya Clement and David Tchong used a video clip of Larry Aitken, an Ojibwe Traditional Knowledge Keeper, to demonstrate how ARLO can be used to identify different kinds of sound in the video– the drum beating, Larry speaking Ojibwe, and Larry speaking English. I have added another set of tags to try to locate when the spirit of the drum speaks and when Larry Aitken answers the drum back to try to reveal the spiritual dimensions of digital technology.

**Kristen Saugee-Beaudry** *Ethical Considerations in Designing Digital Environments with Indigenous Communities*

Traditional knowledge dissemination in Cherokee culture differs from practices that privilege artistic and intellectual freedom. Incorporating Cherokee consultants into all phases of design and development of online Cherokee knowledge repositories is crucial to successful and ethical scholarship.

3:15 pm –

5:00 pm

**Panel 4**

5:00 pm –

?? pm

El Mercado Happy Hour

**Day 2**

8:30 am –

9:00 am

Continental Breakfast

9:00 am – 9:15 am	<b>Day 2 Introduction:</b> Tanya Clement Where do we go next: presentation on what is the follow-up project?
	<b>Panel 5</b>
9:15 am – 10:45 am	<b>Classification:</b> feature discovery, the iterative approach, and looking at features in context (David Enstrom) <b>Clustering</b> (Elizabeth Roke, Chris Mustazza, Michael Nardone)
10:30 am – 10:45 am	Break
10:45 am – 11:30 am	<b>PennSound Clusters:</b> David Tcheng
11:30 am – 12:15 pm	<b>Workflows</b>
12:15 pm – 1:45 pm	<b>Working Lunch:</b> Celebrating the StoryCorps / Benson Collection partnership (Box Lunches, Tocker Lounge, 1st floor)
1:45 pm – 2:45 pm	<b>Planning the final recommendations:</b> small groups (individual google docs)
2:45 pm – 3:30 pm	<b>Planning the final recommendations:</b> larger groups
3:30 pm – 3:45 pm	Break
3:45 pm – 4:30 pm	<b>Planning the final recommendations:</b> individual project write-ups
4:30 pm – 5:00 pm	<b>Final discussion and concluding remarks</b>
6:00	Dinner at the Clay Pit (across the street from the iSchool)

## **Application**

### **I. Title**

Analyzing High Performance Sound Technologies for Analysis and Scholarship

### **II. Investigators (co-investigators)**

Tanya Clement

### **III. Hypothesis, Research Questions, or Goals of the Project**

The goal of this project is to observe and better understand the development and use of computational tools for software developers, librarians and archivists, and humanities researchers who work with sound collections. The hypothesis is that humanists interested in sound scholarship, stewards of sound collections, and computer scientists and technologists versed in computational analytics and visualizations of sound will develop more productive tools for advancing scholarship in the humanities in spoken text audio if they learn together about current practices, if together they create new scholarship, and if they consider the needs, resources, and possibilities of developing a digital infrastructure for the study of sound together.

### **IV. Background and Significance:**

There are hundreds of thousands of hours of important spoken text audio files, dating back to the nineteenth century and up to the present day. Many of these audio files, which comprise poetry readings, interviews of folk musicians, artisans, and storytellers, and stories by elders from tribal communities contain the only recordings of significant literary figures and bygone oral traditions. These artifacts are only marginally accessible for listening and almost completely inaccessible for new forms of analysis and instruction in the digital age. For example, an Ezra Pound scholar who visits PennSound online and would like to analyze how Pound's cadence shifts across his 1939 Harvard Vocation Readings, his wartime radio speeches and his post-war Caedmon Recordings (June 1958) must listen to each file, one-by-one, in order to establish a look at how (or if) patterns change across the collection. An Ojibwe oshkabewis ("one empowered to translate between the spiritual and mundane worlds") seeking to teach students about the ways in which an Ojibwe elder uses Ojibwemowin ('the Ojibwe language') at culturally significant moments to enhance English descriptions with spiritual elements has few means to map or show students when these transitions or "traditional cultural expressions" (TCE) occur. And a scholar doing research within the Oral History of the Texas Oil Industry Records at the Doph Briscoe Center for American History can only discover the hidden recording of Robert Frost reading "Stopping by Woods on a Snowy Evening" among other poems on Side B. of folklorist William A. Owens' recordings because a diligent archivist included that fact in the metadata.

Not only do scholars have limited access to spoken word audio, but their ability to do new kinds of research (what Jerome McGann calls "imagining what you don't know") and to share these methodologies with colleagues and students is almost entirely inhibited by present modes of access. What other TCE's and important historical moments are hidden in these sound files? What if we could test hypotheses concerning the prosodic patterns of beat poets in comparison to the "high modernists" with over thirty-five thousand audio recordings in PennSound? What if we could automatically detect the difference between poetry and prose to determine when a poem is over and an author is telling us about the poem? Or determine, perhaps, whether and when the Ojibwe storytellers sound like elders from supposedly unrelated tribes? At this time, even though we have digitized hundreds of thousands of hours of culturally significant audio artifacts and have developed increasingly sophisticated systems for computational analysis of sound, there is no provision for any kind of analysis that lets

one discover, for instance, how prosodic features change over time and space or how tones differ between groups of individuals and types of speech, or how one poet or storyteller's cadence might be influenced by or reflected in another's. There is no provision for scholars interested in spoken texts such as speeches, stories, and poetry to use or to understand how to use high performance technologies for analyzing sound.

In response to this lack, this researcher Tanya Clement at the School of Information (iSchool) at the University of Texas at Austin (UT) and other researchers at the Illinois Informatics Institute (I3) at the University of Illinois at Urbana-Champaign (UIUC) alongside humanities scholars and stewards (librarians and archivists) are collaborating to develop high performance sound technologies for analysis and scholarship. My study will be to observe and interview these collaborators (which include software developers, librarians and archivists, and humanist researchers) to gauge how their work together influences the development of these tools. As well, I will also observe, interview, and survey users of the tools before and after their interactions to determine how these tools affect their research practices.

A software development team is currently forming at UIUC to work with technologists, librarians, and humanities scholars at PennSound (<http://writing.upenn.edu/pennsound/>) to start developing a system for looking at the 100,000 hours of digital poetry files currently housed there. At UIUC, the "Structural Analysis of Large Amounts of Music" (SALAMI) project has developed audio segmentation tools in NEMA (Networked Environment for Music Analysis) that reveal repetitive structures in music such as "chorus" and "verse" or larger elements like movements in symphony. The SALAMI PI, Stephen Downie, and the developers from I3 who helped develop NEMA are possible people to interview for this study since the work of SALAMI is being repurposed for developing new software to analyze sound files).

Other possible participants for interviewing include stewards of the audio collections of the Ojibwe people or of folklorists John A. Lomax, William A. Owens, John Henry Faulk, Americo Paredes, and Mody Boatright at the Briscoe Center. These participants will be interviewed to gain insight into their current practices and how they might use such a system once developed. Not learning to use these resources from multiple disciplines could mean losing them and losing sight of the kinds of questions different perspectives such as those represented by working poets and folklorists or communities such as the American Indian Library Association could afford. In August 2010, the Council on Library and Information Resources and the Library of Congress issued a report titled *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age*. This report suggests that if scholars and students do not use sound archives, our cultural heritage institutions will be less inclined to preserve them. As a result, archives and libraries must collaborate with patrons and scholars to understand how recordings might be used in the future (16). Nancy Davenport, then president of CLIR, surveyed scholars whose work is primarily with audio and concluded that scholars wanted unfettered access and better discovery tools for what she calls "deep listening" or "listening for content, in note, performance, mood, texture, and technology" (41; 157). Finally, the report suggests that training for archivists and librarians in sound preservation must include "critical listening" skills and "relevant experiences in ethnomusicology, oral history, radio or music" (147).

Beyond computer scientists and the stewards of sound collections, I will interview humanities scholars who work with sound files. If librarians and archivists need to know what scholars and students want to do with sound artifacts in order to make these collections more accessible, then humanities scholars, arguably, also need to know what kinds of analysis are possible in an age of large, freely available collections and advanced computational analysis.

Assuredly, computer scientists and technologists developing systems to allow users better access to oral traditions in folk, poetry, and tribal collections will also benefit from the relevant experiences of scholars and students who already work with these sound artifacts. Charles Bernstein, Donald T. Regan Professor of English and Comparative Literature at the University of Pennsylvania and the Director of PennSound calls literary scholarly inquiry into sound (or critical listening) “close listening;” he closely identifies that activity with increased access, in which case “the sound file would become . . . a text for study, much like the visual document. The acoustic experience of listening to the poem would begin to compete with the visual experience of reading the poem” (*Attack of the Difficult Poems* 114). Humanist scholars involved with the PennSound project, the American Philosophical Society, the Doph Briscoe Center, and the Lyndon B. Johnson Library will also form a subset of the participants I will observe and interview.

Questions that I will consider as I observe the development of these new technologies is how infrastructure development for analyzing sound impacts how and what we learn about cultures. Computer performance, in terms of speed and storage capacity, has increased to the point where it is now possible to analyze large audio collections with high performance systems. As a result, I will consider if developing systems based on the “close listening” practices that literary scholars and scholars interested in folk and tribal collections employ alongside the “distant listening” practices that automated discovery, classification and visualization encourage are critically productive tools that scholars and instructors will want to use.

**V. Research Method, Design, and Proposed Statistical Analysis:**

I will conduct an ethnographical study with observations at meetings with the larger development group including developers who are developing the software, with stewards such as librarians and archivists who are managing sound collections, and with users who are using the software. I will also interview these developers, librarians and archivists, and users individually about the process of development and use. I will finally observe, interview, and survey users I will promote through professional conferences and the institutions named above who are involved in developing new software for these investigations into sound.

**VI. Human Subject Interactions**

**A.** Participants include computer scientists, stewards (librarians and archivists) of cultural heritage sound collections, or humanities researchers who analyze sound. These participants fall into two groups. They are either part of a team that is developing software to analyze sound files or they are potential users of such software. Potential participants include computer scientists at the Illinois Informatics Institute (I3) at UIUC and librarians and archivists and humanities researchers at PennSound, the American Philosophical Association, the Doph Briscoe Center, and the Lyndon B. Johnson Library, all of whom are actively engaged in developing software for the analysis of sound in the humanities. Other participants will be researchers promoted through professional organizations and these institutions. The subject population will include between 40 and 150 participants, all over 18 years of age. They will be included based on their participation with the above mentioned institutions and their research interests in collecting or analyzing sound files in the humanities. These participants will be promoted at professional conferences and through the abovementioned institutions. Participants will be observed, interviewed, and surveyed over the course of the project. All of the interviews and surveys will be conducted in English.

**B.** Beyond the core development group, I will attract a diverse range of participants by advertising this study at each of the collecting institutions including PennSound, the American Philosophical Association, the Doph Briscoe Center, and the Lyndon B. Johnson Library. I will disseminate the call through prominent digital humanities networks such as the Humanist listserv and Twitter. I will also advertise through the National Poetry Foundation, the Society of Southwest Archivists, which will be convening in Austin, May 29-31, 2013. As well, I will advertise the Institute through a variety of groups associated with the American Indian Library Association (AILA), the Association of Tribal Libraries, Archives and Museums (ATALM) and the International Federation of Library Associations and Institutions (IFLA) SIG on Indigenous Matters. Melissa Pond, Director of Library Services at Leech Lake Tribal College and a collaborator with the APS Native American Projects and Sandy Littletree of the AILA have agreed to describe this study to the potential participants and ask them to contact me if they are interested in talking to me about the study.

**C. Describe the procedure for obtaining informed consent.**

For the development group, I will obtain informed consent by providing a consent form to participants and having them provide verbal (not written) consent. In the case that they agree to the audio recording of their interview that consent will be recorded at the start of the recording. In the case of surveys, the first page of the web survey will include an informed consent form. In the case of observations, participants will be given a consent form at the beginning of the observation.

My interviews and surveys include questions that are a normal part of professional conversations, such as reasons for decisions to use or develop a particular piece of software. Many of my interviews will be conducted by phone and/or online conferencing and, in the case of the survey, a web survey.

Observations will take place in the participant's place of work.

Please see the consent form appended to this proposal.

**D. Research Protocol.**

I am asking participants to participate in interviews ranging from 30 to 120 minutes (typically 60 minutes) to discuss their work with sound files and software in humanities research. The interviews are semi-structured.

I am also observing archivists and digital humanists at work. Typically, each person will be observed three times in two or three-hour time-periods.

Please see appended interview and observation protocols.

**E. How will you protect the privacy and confidentiality of participants?**

The privacy of the participants is protected by the process of informed consent and the ability to withdraw from the study at any time.

The confidentiality of the participants will be protected (unless participants give written request otherwise), by not mentioning their names in publications or presentations.

**F. Discuss the procedures that will be used to maintain the confidentiality of the research data.**



Interview recordings will be held confidential to the researcher on encrypted UT computers and/or in locked UT offices. The recordings do not contain the participants' names. As stated in the consent form, these recordings will be stored for a period of five years for use in qualitative analyses, ensuring that the perspectives of the participants are available to inform theory and on-going research.

**G.** Please describe your research resources.

The research takes place primarily via calls between the UT office of the PI and the research subjects, either by phone or online conferencing (e.g., Skype). From time to time interviews might take place at conferences or workshops. Observations will take place in the participant's place of work.

**VII.** Describe any **potential risks**

The study poses minimal risk for the subjects. The primary risk is that interviewees might identify shortcomings in current collection practices or software development in the humanities that could potentially be sensitive in their workplace by identifying places for improvement in their own practices and at their home institutions.

**VIII.** Describe and assess the **potential benefits** to be gained by participants (if any) and the benefits that may accrue to society in general as a result of the planned work. Discuss the risks in relation to the anticipated benefits to the participants and to society.

The participants will be explaining their choices and attitudes about collection and software development with sound files used in humanities research. Intangible benefits include a chance to reflect on their practices and to understand better the implications of architectural and sourcing decisions on long-term maintenance costs for their archival, software development, and research work.

**IX.** Indicate the specific **sites or agencies involved in the research project** besides The University of Texas at Austin.  
None.

**X.** If the project has had or will receive **review by another IRB**, indicate this.  
None

## Works Cited

Bernstein, Charles. "Introduction." In *Close Listening: Poetry and the Performed Word*. Oxford University Press, 1998: 3-26. Print.

Council on Library and Information Resources and the Library of Congress. "Sound Recording Collections: An Overview of Preservation and Public Access in the Twenty-first Century." In *The State of Recorded Sound Preservation in the United States: A National Legacy atRisk in the Digital Age*. National Recording Preservation Board of the Library of Congress. 29 Aug. 2010: 9-65. Accessed 24 Feb. 2012.  
<http://www.clir.org/pubs/reports/pub148/reports/pub148/pub148.pdf>

## Interview Questions for HiPSTAS: March/April 2013

### Analyzing High Performance Sound Technologies for Analysis and Scholarship Interview Protocol: March/April 2013

#### **Introductory Statement, to establish “ground rules”**

1. confirm that the participant has read and understands the consent form
2. confirm that they give permission for the interview to be recorded.
3. in particular confirm with the participant that it is permissible for the interview to take place in their workplace (if that’s where you are).
4. In particular also confirm that we’re not looking for personal, private information or information which could be harmful, and that they should be particularly aware of such risks when we ask about their perceptions and understandings of *other people*.

(begin recording)

#### **Again confirm, on tape, that:**

5. they give permission for the interview to be recorded,
6. they understand the informed consent that they signed
7. it is permissible to hold the interview in the workplace
8. that they should not reveal potentially harmful information about themselves or others.

#### **Introduction:**

“Thanks for taking the time to speak with us today. We sincerely appreciate your time and energy in assisting us with gathering insights into your work with sound files and the digital humanities.”

#### **Articulate objective of interview:**

“We hope to glean user requirements for the ARLO system based on insights and understandings into how you, a digital humanities scholar, are currently working with high-performance sound files.”

**General Themes of questions:** need more info to answer well; will need to map the question-themes to the use cases and requirements documentation (move from general to specific; from familiar to unfamiliar).

#### **Basic Information:**

1. What is your disciplinary or professional area of expertise?
2. What is your experience with creating access to or doing research with sound collections?
3. With which collections are you most interested in working during the course of this Institute?
4. What are the possible areas of study or disciplinary backgrounds of users who would typically also use these collections?
5. What kinds of questions do you anticipate users might want to ask when accessing, using, or sharing these collections?
6. This institute proposes to enhance access and analysis to spoken text collections. How would you define access to these collections? What are we accessing? How would you define analysis with these collections? What is the contribution to knowledge a user might make?

**In general, sound technologies:**

7. In general, how can access to your collection of interest be enhanced by infrastructure that supports advanced means of access?
8. In general, how can access to your collection of interest be enhanced by infrastructure that supports advanced means of analysis?
9. In general, what kinds of access or scholarship do you anticipate will be easy to enhance? Why?
10. In general, what kinds of access or analysis do you anticipate will be difficult to enhance? Why?

**In specific, sound technologies:**

11. How do you currently access or perform analysis with your collections? Specifically, what factors of sound do you explore?
12. What would you like to do that you cannot currently do? When has your current means of exploration not met your needs? Specifically, what factors of sound would you like to explore?
13. What do you anticipate are the specific hypotheses you will explore with your collection of interest in this project?
14. Is this similar to or different than what you do currently?

**In conclusion:**

15. Do my questions seem relevant? Is there anything else I should be asking?

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
ID	Category	Sub-Category	Description
3	Consistency	Metrics	Metrics (between 0 and 1) for classification and clustering is different
4	Consistency	Terminology	consistent naming (spacing and capitalization) of: TagSet, TagClass, MediaFile
5	Consistency	Terminology	cluster tags are new ids confusing when they were clusters of tags that already had an id
6	Consistency	Terminology	In cluster searches, consistency in ID numbers, and consistency in tag strength
7	Design	Architecture	improve SupervisedTagDiscovery addTagExample to DB performance
8	Design	Architecture	Sharing of projects
9	Design	Architecture	search interface for metadata so you can see who else has uploaded files with same types of metadata; store more metadata about actual files we're ingesting; title,
10	Design	Architecture	Create a permanent shareable link to a file
11	Design	Architecture	extended user settings infrastructure
12	Design	Architecture	change queueing system so job with lots of files doesn't take over
13	Design	Architecture	Integration of other feature extraction tools, like FFT
14	Design	Architecture	Integration of other tools: a. Weka capability for more modeling capabilities
15	Design	Architecture	Sharing of results via persistent link (RestAPI call): a. Ability to share catalog views with other users and to embed in publications

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
16	Design	Architecture	<del>Site integration (very ambitious, long term ...):</del> Allow for an ARLO interface to be embedded within a site, exposing its file repository, making that audio archive available for user access
17	Design	Data	import m4a
18	Design	Data	Mass File Upload - Web interface to select and upload a number of files at once.
19	Design	Data	export results as sound clips
20	Design	Data	Being able to export a sequence of machine tags audio-wise
21	Design	Data	download batch of matching tagged clips as a set, or extract the data for the set, from the catalog report of tag matches
22	Design	Data	provide cluster similarity
23	Design	Data	random samples: save sets of random samples, option to tag automatically your random sample
24	Design	Data	Cluster management: Following an unsupervised clustering, it is now possible to delete / accept individual tagged elements. Request: ability to "copy" a class of
25	Design	Data	export the results as sound clips for sharing
26	Design	Data	takes in additional metadata already associated with the file; Overall, there's a lack of descriptive metadata. In addition, there's no space for metadata about the files
27	Design	Data	globus for file transfer
28	Design	Data	change sharing
29	Design	Data	add meta data (could be imported from BWF files)
30	Design	Data	Uploading files can be difficult--size not a problem: a few are stuck in the same blank-white-field-with-orange-line. Also needed: to retain "native" sampling rate at time of

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
31	Design	Data	modify the export for readability and/or usability
32	Design	Data	Have a master library for all the files and be able to create a subset for individual testing. Folks want to be able to create subsets of libraries to run smaller tests. This is
33	Design	Data	Import MP4s
34	Design	Data	Custom Metadata (Long Term)
35	Design	Data	Default for uploading MP3 and converting them to WAV should be that the "native" sampling rate is retained
36	Design	Data	Create Library from Existing Files Select a subset of files from an existing Library to create a separate library.
37	Design	Data	Metadata -- import, output: to be able to subset your data based on it; identification
38	Design	Data	Library/Project sharing - Need to be able to compare across projects, but have data separated into multiple projects; duplicate files (sharing of same file)
39	Design	Data	timestamp metadata (that includes relative location to the entire file); a little bar above or beneath each tag
40	Design	GUI	cluster a window size of file for all my files with or without overlap
41	Design	GUI	multiple comparisons, See multi-file view for comparison (Playlist view?)
42	Design	GUI	Usability: Workflow of tool usage

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
43	Design	GUI	Libraries>Add Audio Files to Library--Some search box or way to add files to library when there's a large collection (right now its a tiny box you can only scroll through)
44	Design	GUI	Context: Enable display/audition of time immediately before and after the selected sample
45	Design	Permissions	perms checking in deleteAllInvalidTagExamples
46	Design	Tags	supervised tag discovery screen -- tag match quality parameter
47	Design	Tags	default your tag sets as not shared
48	Design	Tags	ability to share tag set with chosen recipients
49	Design	Tags	ability to let people read or write your tag set
50	Design	Tags	ability to "unshare" your tag set
51	Design	Tags	autorefresh tagcounts on projectLibraryFiles page
52	Design	Tags	Need to automatically generate 1 sec clips for entire file
53	Design	Tags	Need to automatically generate random samples across the file
54	Design	Tags	save sets of random samples
55	Design	Tags	Move tags between TagSets



## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
56	Design	Tags	visually see hierarchy of tags/tagsets
57	Design	Tags	Malleable tag start/ends--drag ends instead of view tag->change numerical end values
58	Design	Tags	Import/export tags
59	Design	Tags	Color coding to show strength of tag matches. Instead of just letting the algorithm show us the outliers, can we also get a visualization of the match strengths? As in
60	Design	Tags	ability to label tagid to provide additional information about what is the tag's significance
61	Design	Tags	Merge Tag Sets
62	Design	Tags	Longer tag class names;
63	Design	Tags	Unique Identifier Naming ability of tag class instances
64	Design	Tags	Ability to create random window tags (what size and from a subset of a library)
65	Design	Tags	Ability to configure tagging interface
66	Design	Tags	improvements for parameter settings - choose a few of these, not all of them: i. voice (male / female) ii. timbre
67	Design	Tags	timeline of tags per file; multiple files at once.
68	Design	Tags	more pure charts that need to be rendered online: i. num tags per strength
69	Design	Tags	tagging utterances (Tony will make fundamental available for Folklore): Enable generation of a sequence of automatic, sequentially segmented tags based on silence and intonational patterns; with the ability for the user to modify and improve

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
70	Design	Tags	Clustering: a. add name label of tagsets to parameters display
71	Design	Tags	Validation <b>access to audio (making it easier to hear what you've tagged):</b> a. Implement auditing of tags without toggling to another window.
72	Design	Tags	Master tag sets ( <b>very ambitious, long term ...</b> ): one for applause or silence, etc., or dog barking, etc.
73	Design	Visualization	matching that was dependent on the way the wave looked that didn't require match on the frequency? Search for wave shape at different frequencies and match those?
74	Design	Visualization	better to start with broad view of whole track
75	Design	Visualization	multiple file view for comparisons
76	Design	Visualization	need label of filename at top of page that displays the vis
77	Design	Visualization	Real time scrubbing and auto-scrolling of audio
78	Design	Visualization	Click to play on spectra and select area to play
79	Design	Visualization	Horizontal Scroll while Playing Audio: When playing a long audio spectra, we should enable automatic scrolling to the side
80	Design	Visualization	Scroll bar to zoom in or out on audio
81	Design	Visualization	Standard waveform superimposed over or displayed under the spectral display
82	Design	Visualization	Instead of having to adjust the time parameters to get a big screen, Zoom In and then Half Back, Zoom In and Half Forward, and then work on Gain, Dampening, and other
83	Design	Visualization	Playback controls everywhere
84	Design	Visualization	Numerical values for the colors we see; also ability to see finer gradients in color scheme (instead of white being 90-100, see what is 95-100)
85	Design	Visualization	On VisualizeProjectFile tags only highlight if the TagSet is owned by the user; other user's tags aren't showing up on the spectrogram

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
86	Design	Visualization	A way to highlight a specific region (highest section of frequency range or lowest section) and hear just that sound instead of the entire file--eg, if someone dropped a
87	Design	Visualization	Put smaller time units on horizontal line of long visualizations. Right now I just have 1 second, 100 seconds. Would like to be able to see down to smaller units of time. Add
88	Design	Visualization	Ability to create new tag sets on visualization screen instead of exiting
89	Design	Visualization	Add minutes to the time-coded fields.
90	Design	Visualization	Add click and drag scroll function.
91	Design	Visualization	Ability within visualization to cut and drag or move a segment to a new window.
92	Design	Visualization	Customizable color spectrum for audio frequency display, especially for displays/presentations
93	Design	Visualization	The ability to edit different colors for different tags
94	Design	Visualization	have the image be more dynamic, so you can adjust some of the levels on a sliding scale and see in real time what happens to the image as you slide through values (e.g.
95	Design	Visualization	when visualizing a file with a setting that is longer than the file length, the audio player is inaccurate
96	Design	Visualization	Configure the "Show PitchTrace" option so that it is possible to visualize the pitch trace without being superimposed over the Spectra.
97	Design	Visualization	Ability to select views of visualization by tag set or tag author. The ability to see shared files with AND without user tags.
98	Design	Visualization	output the visualizations

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
99	Design	Visualization	Pitch tracing: a. Add the option for "true" pitch-tracing
100	Design	Visualization	Predictive Modeling: integrate visualization tools for interpreting confidence rates for match strength
101	Efficiency	Shortcuts	seems like too many clicks to get to the tagging step
102	Efficiency	Shortcuts	Project links at top: dropdown Libraries and catalog w/o having to open project page
103	Efficiency	Shortcuts	default settings

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
104	Efficiency	Shortcuts	Save different settings for different log-in events
105	Efficiency	Shortcuts	When choosing a tagSet to display, keep that selection
106	Efficiency	Shortcuts	allow multiple user settings
107	Efficiency	Shortcuts	default tagging mode (adv / basic) a user setting
108	Efficiency	Shortcuts	User customizable pages - so features not desired do not need to be listed for each user. Perhaps have a settings page with the boxes to select or unselect those features

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
109	Efficiency	Shortcuts	When going to Library page, currently have to press submit to get audio to display.
110	Efficiency	Shortcuts	On the catalog page, only display the tag classes you want instead of them all.
111	Efficiency	Tags	unaccept a tag you have accepted
112	Efficiency	Tags	on RandomWindowTagging screens, default the create tag form to the values of the current tag window, so 'create' will just create the whole window as a tag
113	Efficiency	Tags	We can individually rename a tag set or cluster. It would be useful to be able to batch rename. Use case: I do an unsupervised search - it produces 100 clusters. cluster01,
114	Efficiency	Tags	Change Tag Set Name, Tag Class Name, and Display Name: On the Visualizer we see the red machine tagged box. We would like to edit this box. We machine supervised
115	Efficiency	Terminology	change TagClass from tag edit form
116	Feedback	System Logs	jobMessages - log info about jobs
117	Feedback	System Status	ARLO should tell you when it's down
118	Feedback	System Status	Status indicators
119	Feedback	System Status	status page is up to date and easy to read -- change status showing on all pages (Job Management Page) (visualize pages)
120	Feedback	System Status	some way to know when your job is submitted and running
121	Feedback	System Status	Show for each user how many jobs are queued in front of them so they have a sense of the wait time

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
122	Help	Documentation	When we ingest files, we want to include descriptive metadata (date, description, content notes, etc.). How should our descriptive metadata match up against existing standards? We don't want to discard existing standards, but is there a way we can
123	Help	Documentation	generate audio thumbnail for a particular speaker based on known examples
124	Help	Documentation	best practices for tagging
125	Help	Documentation	example projects that identify different things
126	Help	Documentation	Update wiki and image for the Project / Library / TagSet hierarchy
127	Help	Documentation	define with examples - visualize hierarchies: projects, libraries, tagsets, tags classes, tags
128	Help	Documentation	Project, Library, Collection, Workset, TagSets, TagClasses
129	Help	Documentation	Should create an image of the current hierarchy for clear discussion.
130	Help	Documentation	add to glossary the terms on the vis page, like gain, damping, etc
131	Help	Documentation	In general, it would be helpful to have more documentation related to what David was explaining to us, so we can more successfully understand the parameters. For
132	Help	Documentation	More clarity on the taxonomy of tagging would be helpful. What are the best practices for tagging in shared system? What is the best strategy for tagging given multiple goals? This is a broader digital humanities question: how can we standardize

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
133	Help	Documentation	What happens if multiple people tag a file?
134	Help	Documentation	readings/bibliography
135	Memory	GUI	We get kind of lost in terms of the hierarchies of our files--maybe we have to get our sea legs, but for those of us who are much more visual learners and not so good at
136	Metaphor	Terminology	Unsupervised Tag Discovery - replace with GetOrAddTagClass()
137	Metaphor	Terminology	in forms, rename 'className' to TagClass
138	Metaphor	Terminology	Metadata on upload media file: still very customized for bird--should have a chance to explore metadata fields that apply better to spoken word files (things like language
139	Metaphor	Terminology	create transcript is the best name/label
140	Navigation	GUI	seems like most functions are on the library page
141	Navigation	GUI	put current job at top, with old at bottom
142	Navigation	GUI	manage jobs instead of catalog could be brought up when job submitted
143	Navigation	GUI	Two or more tag searching --Aggregate tags to search / A then B search
144	Navigation	GUI	Move match quality parameter from catalog to supervised tag discovery
145	Navigation	GUI	simple playback everywhere (e.g. catalog)
146	Navigation	GUI	link to adjust userSettings on every applicable page



## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
147	Navigation	GUI	Cluster displays: allow folks to sort worst clusters to the top or somehow get to the middle. Show the cluster confidence. Allow Play all the clusters at once all across in the catalog. allow the user to decide how much time a segment or window size and overlapping is for clustering.
148	Navigation	GUI	in adding media files to library user needs better way to find media files through searching or sorting or browsing, etc.
149	Prevention	Error	When attempting Supervised Tag Discovery, I chose two discovery classes to run simultaneously. The attached image shows the "Manage Jobs" Display, which said
150	Prevention	Error	no way to delete projects. Doesn't work.
151	Prevention	Error	For SupervisedTagDiscovery (non-QueueRunner), upon completion isRunning and isComplete is not updated.
152	Prevention	Error	empty tags (classes) should not be shown in the catalog (at least by default)
153	Prevention	Error	When deleting a TagClass in the catalog, the TagClass still shows up on the next refresh, then creates an error since the TagClass is still in the select box, and creates a validation error on the next refresh.
154	Prevention	Error	save settings from run to run, create new settings doesn't work
155	Prevention	Error	no way to detect other usernames; is there a way to do this? (via dynamic lookup list for group?)

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
156	Prevention	Error	radio button for exchange or move mode (not checkbox)
157	Prevention	Error	Don't allow users to set frequency rates that are impossible to detect beyond the nyquist rate
158	Prevention	Error	When deleting an audio File from the media files page, the response page is the delete URL - this breaks refresh
159	Prevention	Error	Move "Delete" link physically away from "View" link
160	Prevention	Error	interface: makes you think classname, b/c not a dropdown, looks like it's a free text, instead of associating with the class that already exists; maybe a warning if a class already exists and you're trying to create something with same name
161	Prevention	Error	no uploading the same file
162	Prevention	Error	Support for unicode characters in metadata, including Navajo, Cherokee and Spanish characters. <a href="http://symbolcodes.tlt.psu.edu/bylanguage/index.html">http://symbolcodes.tlt.psu.edu/bylanguage/index.html</a>
163	Prevention	Error	restrict to only delete from your own tagsets

## ARLO Requirements (Open)

1		Visualization	visualize hierarchies: projects, libraries, tagsets, tags classes, tags
164	Prevention	Error	Bug Fixes/Improvements: a. match strength was 1.003 (greater than one)
165	Prevention	Forms	Highlight required fields on forms (LOE 2 to 8 hours)
166	Prevention	Forms	Upload MediaFile page - when a project is shared w/ multiple users, the Project dropdown gets overpopulated
167	Recovery	Error	deactivate all audio files generates an error
168	Recovery	Forms	Forgot password reset form
169		Documentation	<a href="http://bigd.ncsa.illinois.edu:81/doc/wiki/index.php/Project_Hierarchy">http://bigd.ncsa.illinois.edu:81/doc/wiki/index.php/Project_Hierarchy</a>
170			Ability to configure validation interfaces

## **AppendixE: Evaluations**

### **Group 1: Archivists**

#### **1. In general, what are the ideal tools that would be most valuable to your community?**

- ability to combine clusters
- ability to drag and drop w/in clusters
- use combined clusters to start supervised tagging
- ability to create simplified input forms for non-expert users
- need additive tagging/tagsets--so having a single “master” tagset for all types of files, plus specialized tagsets for different types of recordings (like an oral history tagset, a music tagset, etc) that can be added to files tagged with the master tagset
- what are the main access points that archivists need to be able to serve out to researchers?  
(dialect/accent--maybe also as a means of identifying geographic area, sex of speaker, genre, etc)
- can we get mechanical/recording device/playback “noise” technical metadata from these recordings through ARLO?
- ID individual speakers?
- speech-to-text capabilities
- administrative tools for things like batch ingesting, permissions controls, etc
- ability for researchers to get results exported and delivered automatically (via email, ftp, or?)

#### **2. In particular, how can a tool like ARLO accommodate current workflow practices?**

- incorporate ARLO at time of digitization to spit out metadata
- or maybe incorporate ARLO once you have a digitized collection rather than at the individual file level?
- shared libraries in order to improve machine learning?
- do we need to have different libraries for different types of recordings? Would it be useful to have one giant library and have ARLO look for commonalities across all types of recordings?
- need additive tagging/tagsets--so having a single “master” tagset for all types of files, plus specialized tagsets for different types of recordings (like an oral history tagset, a music tagset, etc) that can be added to files tagged with the master tagset
- also be able to incorporate researcher or content-specialist tagsets?

#### **3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?**

- timestamping (at least as a CSV), output with the tag
- ability to display in a DAMS or CMS (this would work if exporting as CSV)
- right now everything is at file level, but would be interesting to pull back and do analysis at library level
- comparing similarity of centroids

**4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?**

--timestamps

--export as CSV, ability to import/ingest external metadata into ARLO

**5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?**

--standardize metadata!

**6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?**

--improved documentation

--documentation of things that ARLO can do--a “how-to” manual for specific research scenarios (a user manual)

--policy recommendations on how we allow researchers to use digital collections for “big data” projects

--naming conventions!

--UI improvements

--tracking provenance of audio snippets during discovery

**7. What are future grant or collaboration opportunities the group might engage to support this work?**

--metadata structures for capturing ARLO’s results outputs

--researcher interface design

--need additive tagging/tagsets--so having a single “master” tagset for all types of files, plus specialized tagsets for different types of recordings (like an oral history tagset, a music tagset, etc) that can be added to files tagged with the master tagset

--set up individual committees/task forces for particular topics--education and training, usability, etc--to evangelize to different groups

--more likely to get more involvement if there is a tangible product that goes back to participating institution?

--support for dissertation fellowship to use ARLO for a dissertation and give feedback on interface design etc

### **Group 3: Native American Sound Archivists, Librarians, and Scholars**

#### **1. In general, what are the ideal tools that would be most valuable to your community?**

ARLO's power seems to be in sound recognition and pattern grouping. We have not really spent time reading the images produced by ARLO.

- We take the word "community" very seriously. Right now, Arlo is very abstract and, we feel, too far removed from what would serve Native communities. What we would like to see is far more practical outcomes. How can Arlo be used for language preservation / revitalization? Because ARLO works better with smaller bits of information, we hope that it could be used to isolate Cherokee vowel sounds that could be used by language students. It would be much more useful if it could be programmed to identify the following:

- Word recognition tool
- Phrase recognition tool
- Utterance recognition tool
- presentation tools or documentation to present sound processing tools and metadata/library science and analysis
- training tools to help community access archives and computer interfaces
- ARLO asynchronous mooc or video training series to introduce community to ARLO and sound research
- ARLO metadata so that interested researchers can see and understand the kinds of projects that are underway and that have been completed with ARLO.

#### **2. In particular, how can a tool like ARLO accommodate current workflow practices?**

ARLO recognizes patterns in thousands of hours of material. We need meaningful clustering into words, phrases, utterances, for our focus on language.

Speaker recognition settings. We are interested in identifying and recognizing the people involved. Right now, Arlo produces "files" and "tag sets." We would like to see "people" incorporated, more specifically the Native people who told the stories that we are analyzing. This metadata is available from the APS at the item or track level. The question then becomes how to incorporate it back into ARLO. The following would also be helpful:

Shifts in speaker. We want to know when a new speaker takes the floor, takes a turn.

Shifts in language. We want files returned that indicate shifts in language and that tag language use.

#### **3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?**

- We would like to see more culturally specific metadata incorporated and generated. For Native American sound archives, the metadata needs to reflect the following, which are rarely included in LOC headings, though are available from more careful readings of the archival materials:

Project names and descriptions and participants

The metadata should include the speakers involved.

- speakers
  - recorders (people who record)
  - language
  - location (community affiliation)
  - dialect
  - tribal or cultural affiliation
  - date of recording
  - family/clan affiliation
  - cultural status (beloved, national treasure, institutionalized roles, tribal historian, traditional knowledge keeper, president, chairman)
  - primary audience (public performance, performance for outside research, performance for inside research, for speakers, elicited, pedagogical, family conversation)
  - genre (personal narrative, new story, old story, conversation, word list, talking to anthropologist)
  - copyright and permissions
  - cultural metadata, cultural importance of digital asset
  - users of data, who is using/research
- (might not fit into ARLO but should be linked to files in ARLO)
- users level of cultural proficiency. What cultural claim does the user make to the material?

**4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?**

- See above
- The metadata should be linked to the chunks produced by automatic recognition across files.  
file name and time in the file linked to the chunks/tags

**5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?**

- funding for technical workers, changing file formats, digitizing, uploading
- data treatment protocols - what happens to the data? Where does it go? How is access and distribution and copying controlled? What are the protections against hacking? When will files be deleted? When will 'a project' be done?
- ARLO should benefit the community somehow. What results can ARLO show that will help people preserve and maintain the languages and cultures and cultural assets? What does ARLO return to the community in return for access to digital cultural assets? To achieve this, tribal communities need to be more involved in the work flow. Lorie suggested that UT might develop an IRB for Native American research.
- assistance in developing digital archives

-cultural training protocols and cultural training modules prior to access to sensitive material

**6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?**

-Letters of support

-hosting ARLO researchers/technicians

-community presentations of research, community discussion of research

-scholars should include native people in the communities, native people should be treated as scholars with respect to their knowledge of culture and history

-communities can outline projects of interest and needs in the community to help channel research curiosity to culturally appropriate and critical research

- help make Digital Humanities scholars more aware of the importance of cultural specificity

**7. What are future grant or collaboration opportunities the group might engage to support this work?**

-Administration of Native Americans technical training grant

-NEH applications

-APS host of workshops, maybe with society of american archivists

-languages of small populations of speakers is a 'hot' topic

-develop next generation of sound researchers in native communities

-Mellon or office of digital humanities grant to bring researchers together into archive to understand sound

-Indiana wax cylinders.

-bringing tribes, archives, and ARLO together

-develop community-based projects involving sound processing and sound processing training that appeal to researchers interested in language, sound, education, visualization, and cultural interaction

-identify critical sound archives that need to be processed, digitized, analyzed, made public

**8. What are future pedagogical or publishing opportunities the group might engage to support this work?**

-ARLO and other visualizers for language learning

-Almost everything we turn up on ARLO is new and interesting

-continuing education modules irb development, tribal laws and ethics, ethics more generally, to deal with issues of digitizing and researching sound archives (families' stories, personal information, cultural information) - data may want to be free but we can control our environment in mutually beneficial and



emotionally satisfying ways. On-line mooc-ish or bricks-and-mortar curriculum and materials development

- Share information about publishing opportunities and appropriate journals
- network with one another and suggest collaborative work and review each other's manuscripts
- advocate for social media as scholarly work, pod-cast, peer-reviewed online HiPSTAS Sound in Digital Humanities journal.

**9. What aspect of the process of developing a tool like ARLO is/was most frustrating? How do we ameliorate?**

- understanding projects and how ARLO can or cannot answer project questions
- cull language material, identify language shifts
- linguistic and technical knowledge background for even starting a project, understanding input parameters, reading visualizations
- zooming and reading visualizations, especially F0, the physical frequency we hear as pitch.

**10. What are more questions we should be asking? (answer them, please)**

- What scholarly research has this project facilitated?
- What are the practical outcomes of Arlo for communities outside of DH?

## Group 2: Poetry Librarians and Scholars

### 1. In general, what are the ideal tools that would be most valuable to your community?

**a. Task-oriented Presets** \*\*\*(for damping, pitch weight etc.: to allow for emphases on pitch, rhythm, bass voice, soprano voice, etc)

*Rationale:* aid in producing more promising initial results than presets, especially for novices.

### **b. Sharing** \*\*

a.) Ability to share catalog views with other users and to **embed in publications** through persistent URLs (see Voyant <http://voyant-tools.org/>)

b.) Means for easily creating temporary or ad-hoc user accounts to facilitate "tagging parties"

c.) Or some form of "**export**" **that allows** you to batch export images / audio files ...

*Rationale:* allow for Hipstas participants to more easily show the results of work and/or to enlist the participation of new users.

### **C. Utterance Tagging/Feature? / Silence Delimiter?** \*\*\*

Enable generation of a sequence of automatic, sequentially segmented tags based on silence and intonational patterns; with the ability for the user to modify and improve them.

(KS: I don't think that a preset can address this; we need the equivalent of the random window generator, but one which segments the file based on the space between them, which probably involves identifying a "background noise" sample for a given file.)

*Rationale:* Auto tagging of random, time-delimited (1, 2 sec.) tags is only more useful for some tasks. Speech analysis may benefit from organic, dynamic sized windows.

### **d. Pitch tracing**

a.) Add the **option** for "true" pitch-tracing--that is, add fundamental frequency option for pitch (autocorrelation), to align with a common approach to measuring and analyzing pitch in acoustic phonetics. And allow users to export images of pitch trace.

b.) Enable the visualization of pitch-tracing outside of the main image windows

*Rationale:* The pitch trace is often difficult to see within the heat-map.

### **e. Context**

Enable display/audition of time immediately before and after the selected sample.

*Rationale:* Will allow for more useful, accurate characterization or validation of short-duration samples.

### **f. Visualization**

a.) Integrate visualization tools for interpreting confidence rates following clustering or supervised tagging.

b.) Integrate visualization of overlapped tagging results (as graph) without heatmap.

*Rationale:* Results can currently be exported and visualized; integrated visualization would allow for more informed decisions in iterative processes.

**g. Audit access (making it easier to hear what you've tagged)**

- a.) Implement auditing of tags without toggling to another window.
- b.) Reviewing clusters, allow for playing of the clustered tags in sequence from one window.

*Rationale:* Ease of use.

**h. Plugin/API/Site integration (very ambitious, long term ...)**

Allow for an ARLO interface to be embedded within a site, exposing its file repository, making that audio archive available for user access. (See TokenX integration into the Willa Cather site as a text- based analog:

[http://jetson.unl.edu:8080/cocoon/tokenxcather/index.html?file=..%2Fxml%2Fcather%2Fcat.ss020\\_sample.xml&fileChooser=on](http://jetson.unl.edu:8080/cocoon/tokenxcather/index.html?file=..%2Fxml%2Fcather%2Fcat.ss020_sample.xml&fileChooser=on))

*Rationale:* There are other desktop tools that do *some* of what ARLO can do. But site integration would increase access and motivate other users / groups to engage with ARLO.

**i. Interface**

We talk about "clustering" but this is called "unsupervised tag discovery" in the interface. These are the same, right? - Correct terminology ...

Draggable tag selection

Ability to adjust machine tags as you make them user tags

***Interface issues: some user optimization.***

- *draggable clip/tag adjustment*
- *audition of tags in sequence, short pause in between*
- *access to visualization of the centroid*

***Group - tagging***

- *access to validation interface*
- *ability to configure a group-tagging interface (akin to the Emory)*

**j. Project organization tools**

More ease in creating subsets of files to work on without creating new projects.

Filter and "Clone" subset of a project.

Also useful to have access to file-based metadata. ;

**k. Automated acquisition and display of ID3 tags from mp3 files.**

**2. In particular, how can a tool like ARLO accommodate current workflow practices?**

(See "sharing" above and "API" above).

Outreach, access, teaching ....

**3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?**

Pennsound uses ID3 so that?

Other known data, such as media-status (born-digital, studio, live, languaget)

Open-field: User input meta-data fields associated with each tagged item (rather than or in addition to just renaming the sample).

**4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?**

See above.

**5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?**

Metadata:

Speaker

Type (reading, speaking)

Location

Date

Gender

Occasion (lecture, performance, ..)

**6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?**

- Faculty need money for course release time to focus on developing collaborative projects.
- Graduate student travel and research; mini-grants?
- Encourage partnerships, say, with SpokenWeb or Editing Modernism in Canada for future collaborations for grant partnerships? (Or other US-based organizations that could fund graduate student involvement with the project.)

-

**7. What are future grant or collaboration opportunities the group might engage to support this work?**

- Grant Funding
- Conferences / Gatherings
- Virtual Meetings

-Prepare a panel for MLA Austin 2016.

-Prepare a seminar panel for an upcoming ACLA.

-Possible gathering in Canada – Montreal? – funded through SpokenWeb or Editing Modernism in Canada research monies. A conference?

**8. What are future pedagogical or publishing opportunities the group might engage to support this work?**

+ DistancedAudioPoetryAnalysis?

Googlesite - to share gray-literature

MM: Maybe we could collaborate on a paper exploring what ARLO has taught us about trends in contemporary poetry performance, with each member focusing on one aspect, e.g., tempo, pitch, rhythm, female vocal range, laughter, etc.

MN: To continue from point above, I think that collaboration on papers is an interesting thing to work out as each one of continues on various researches. Perhaps best way is to create a wordpress we contribute to? One to contribute notes and thoughts on experiments. These would be notes or initial writings that might lead to a more public form of publication in Jacket2 or elsewhere.

**9.What aspect of the process of developing a tool like ARLO is/was most frustrating? How do we ameliorate?**

See above.

**10. What are more questions we should be asking? (answer them, please)**

What's the process for developing interesting research questions, and sharing them? How can we move towards collaborative scholarly questions? Are there ways that we can generate certain general questions. How do we sustain research/collaborative energy once HIPSTAS F2F sessions have transpired?

## **Group 4: Sound Scholars**

### **1. In general, what are the ideal tools that would be most valuable to your community?**

A community focused on using ARLO to look for patterns on music as sound. The language and sound folks seem to want to focus on shorter time windows of spectra and sound analysis. We tend to want longer time windows for musical analysis (a line of a song lyric, a particular melodic phrase, a particular rhythm). The longer time window tends to pull in a lot of sonic information, so then we need better strategies for sorting that out: for instance, options to tag a longer tag and then sort things out within it). Tips, strategies, documentation, defaults, starting points for how to isolate a “unit” of useful sound for further algorithmic (clustering) and visualization analysis when using a longer tagged sound element. Also a better way to keep notes on different results in ARLO as one works iteratively. Is that metadata?

### **2. In particular, how can a tool like ARLO accommodate current workflow practices?**

The question is how to use clustering in service of discerning musical patterns. We need better documentation of what different parameters reveal (for instance, visual examples of damping at different settings with other parameters the same; proposed default settings for voice, different instruments, relations among different sounds, just as starting points to speed up work flow, number of clusters, strategy tips if one is looking for “pure” matches or wants to explore “impure” relationships). The goal is to speed up work flow at the start, so that we can get into more iterative exploration from some starting point that we can grasp as useful. Documentation, examples, tips, strategies, proposals for different research questions would help a lot.

### **3. What ideal data or metadata should a tool like ARLO use or incorporate into analysis?**

### **4. In particular, what ideal data or metadata should a tool like ARLO produce? What would it look like? How would it be organized?**

We are grouping answer to 3 and 4 here. More info on time stamps in results to create metadata about a generalized picture of a track. Right now we tilt toward percentage of strength of match. We want ARLO to help us visualize sequencing in a dataset. For instance, could we visualize the progression of timbre patterns across Vol 1 of Harry Smith’s Anthology of American Folk Music. So that the tool produces not just percentage of strength in matches, but a kind of durational story about sound across a recording.

Can ARLO generate the data/metadata to allow a user to see where similarities are occurring across a dataset. For instance, ARLO returns similarities of pitch interval between different songs on Harry

Smith's Anthology of American Folk Music. I want to be able to start to map these in terms of where Smith located the songs on the Anthology. So a way to extract percentage of strength in mapping to location of those similar sounds within the dataset (in this case the Anthology. So ARLO shows me Fishin' Blues chorus is similar to John Hardy chorus, now I want to be able to visualize where those songs appear on the Anthology to think about how Smith was sequencing them in his sonic mix).

**5. How can current workflow practices at an institutional or community level (yours or in general) change to make data more accessible to a tool like ARLO?**

Building an ARLO API: Working on a more linked open source URL model of uploading sound files and data into ARLO. And vice-versa, can we pull files out of ARLO more easily, expose its data more easily for use.

**6. In general what can cultural institutions or communities do to support these new research activities? What ongoing support is needed to make research and development useful to scholars in the future?**

I think we all need more training in sound studies and thinking through algorithms, so we can conceptualize what ARLO is doing with sound, but also what sound is, how we conceptualize it through frequency, rhythm. Also, how he is thinking about structuring the algorithms. With more knowledge at the ground level of these issues, we can start to use the tool more effectively. But it's also more work. Maybe we can use the documentation and default strategies or some kind of online sound studies/algorithm conversation to further our knowledge about what ARLO is doing with sound and algorithmic models of processing it. : a reading list from David T. on this, interviewing David T. about the conceptualization of tuning forks, frequency, damping, so that people are able to watch and rewatch him explaining the inner workings of ARLO, what it can do, what it can't, \*how\* it is conceptualizing sound, how he is thinking about the design of the algorithms for clustering. We all can use the tool better if we can continue to see what's "under the hood" of its design more clearly. This takes time for each of us to "get it." I know there's a temptation to want to shut the hood and automate this like it's magic. No! Better to educate users here. Maybe video clip interviews with David on how he designed tool...

**7. What are future grant or collaboration opportunities the group might engage to support this work?**

Possible partners:

-Association for Cultural Equity - building on Alan Lomax's work on cantometrics as ancestor to this work. Not just technically, but a paper or study that maps out a deeper history of how ARLO is a descendent of Lomax's Cantometrics/Performance Style, which used computation/clustering. Bring ARLO into the sphere of the Global Jukebox Project that ACE has been pursuing.

- Grammy Foundation
- Association for Recorded Sound Collections
- Rockhall
- Country Music Hall
- Experience Music Project (Paul Allen, MicrosoftConnection)
- Google Cultural Institute
- Apple
- Southern Folklife Collection, UNC Chapel Hill (Steve Weiss)
- Kaiser Foundation (Woody Guthrie Museum)
- Knight Foundation?
- Mellon?
- David and Reva Logan Foundation (funded Jazz Loft Project )

**8. What are future pedagogical or publishing opportunities the group might engage to support this work?**

Appoint someone to be the editor/facilitator and pay that person or give them some kind of reward. A small lowercase journal using ARLO, ARLOResults? Studies with ARLO? ARLOmetrics? Thinking of a name for the journal. Spectralanalysis Studies? These are silly but you get the idea. An online space for registering research with ARLO. Peer review? Different kinds of articles, sound samples, reports, interviews, research questions, how-to guides, a place to catch and preserve ideas about how to use ARLO and results generated from it. Make it agile, iterative publishing, be creative in different kinds of publication models and strategies and levels. Maybe this could be a collaboration with an existing academic journal: History Workshop Online? Amodern? Or a sound studies journal (Viet Erhlman's new one at UT? A co-sponsored online journal b/t music, history, literature, linguistics, indigenous studies, cultural studies, sound studies journals? Have to find the right fit and make the right agreements about control/freedom over the online space, but a collaboration could be productive.

**9. What aspect of the process of developing a tool like ARLO is/was most frustrating? How do we ameliorate?**

**10. What are more questions we should be asking? (answer them, please)**

- A better way into using ARLO through improved documentation and default tips as starting points.
  - An online, multimedia space for publicizing experiments with and research conducted through ARLO
- This becomes a platform for spreading awareness, use, and worthiness of the tool beyond the original HiPSTAS group, but using the work we've done and want to do in service of that expansion.



## CFP

### Call for participation closed.

The HiPSTAS project invites applications for its 2013 NEH-funded Institute for Advanced Topics in Digital Humanities. We encourage a diverse range of librarians, archivists, scholars (including graduate students), and cultural heritage professionals from all types of institutions, disciplinary backgrounds, and expertise, who are interested in working with sound collections and technologies to apply. Members of the American Indian community, in particular, are strongly urged to apply.

### Important information for applying:

1. **Deadline: Extended to February 1**

This is an online application form. After submission, you will receive an email confirmation within 24 hours of receipt.

2. **Meetings:**

The first HiPSTAS meeting will take place in Austin, TX from May 29 – June 1, 2013. The second meeting will be in May 2014, location to be determined.

3. **Participating HiPSTAS collections:**

The collections are described on the [Resources](#) page under “Collections,” but participants are welcome to identify a collection outside of the participating collections with which they will work over the course of the HiPSTAS Institute, such as one from their home institution. [Note: Outside collections must be made available for analysis by HiPSTAS software. Please send questions to [hipstasinfo\[at\]utlists.utexas.edu](mailto:hipstasinfo[at]utlists.utexas.edu) for details].

4. **Advanced means of access and analysis:**

Modes of advanced analysis and access that are available as part of this project are described on the HiPSTAS [Resources](#) page under “Software,” on the [About](#) page, and on the [Meetings](#) page.

5. We will accept two primary types of proposals to encourage both scholars and stewards of sound collections to apply to the HiPSTAS Institute.

#### Contact

[hipstasinfo\[at\]utlists.utexas.edu](mailto:hipstasinfo[at]utlists.utexas.edu)

#### Search

To search, type and hit enter

#### Recent Posts

- [HiPSTAS NEH Institute Final White Paper](#)
- [MLA 2016: Close and Distant Listening to Poetry with HiPSTAS and PennSound](#)
- [John A. Lomax and Folklore Data](#)
- [Hearing the Audience](#)
- [Marit MacArthur receives ACLS digital innovation fellowship](#)

#### [Project Pages](#)

#### [Discussion Forum](#)

#### *Meetings Storified:*

[May 29](#)

[May 30](#)

[May 31](#)

[June 1](#)

#### Meta

- [Site Admin](#)
- [Log out](#)

### **Type one proposals: Access**

Participants who wish to focus primarily on increasing access to sound collections must include a 1000-word description of the ways in which access to their collection of interest would be enhanced by infrastructure that supports advanced means of access to sound collections.

### **Type two proposals: Analysis**

Participants who wish to focus primarily on analyzing sound collections must include a 1000-word proposal that includes a clear description of a possible research question and how this research question pertains to a particular audio collection. Strong proposals will identify a potential audience for this research (such as, but not limited to, folklore studies, literary studies, history, sound studies, or tribal culture studies) and will consider possible connections or comparisons across audio files in one or more collections.

### **6. Commitment expectations:**

Applicants will also be asked to agree to four specific and central HiPSTAS commitments:

- A. Participate in 10 hours of pre-meeting training with the HiPSTAS tools. This training will be virtual and will not require travel.
- B. Attend the first four-day meeting, held at the iSchool at UT-Austin in May 29-June 1, 2013.
- C. In the interim year, meet virtually with the Institute Co-PI's and report periodically on the HiPSTAS blog about use case and ongoing research within the project.
- D. Attend the second meeting, a two-day symposium in May 2014 at UT-Austin (the site may change depending on participant needs).

### **7. Support for participants:**

Travel to and from Austin for both Institute meetings will be funded by HiPSTAS up to \$500 for each meeting. Lodging for 4 nights at the first meeting and for 2 nights at the second meeting will be arranged and paid for by HiPSTAS. Breakfast and lunch will be provided each day.

[Edit](#)

## **4 Trackbacks**

- By [Welcome to HiPSTAS](#) on November 14, 2012 at 9:29 pm | [Edit](#)
- By [Welcome to HiPSTAS](#) on November 16, 2012 at 11:37 pm | [Edit](#)
- By [High Performance Sound Technologies for Access and Scholarship | HiPSTAS](#) on January 8, 2013 at 6:16 pm | [Edit](#)
- By [Deadline Extended: HiPSTAS project « Technē Institute](#) on January 30, 2013 at 6:26 pm | [Edit](#)

## **Post a Comment**

Logged in as [Tanya Clement](#) [Log out?](#)

Comment



You may use these HTML tags and attributes `<a href="" title="">` `<abbr title="">` `<acronym title="">` `<b>` `<blockquote cite="">` `<cite>` `<code>` `<del datetime="">` `<em>` `<i>` `<q cite="">` `<s>` `<strike>` `<strong>`

Post Comment